



Supplementary Information for

Modeling the evolution of the U.S. opioid crisis for national policy development

Tse Yang Lim^{1,2}, Erin J. Stringfellow³, Celia A. Stafford^{3,4}, Catherine DiGennaro³, Jack B. Homer^{1,5}, Wayne Wakeland⁶, Sara L. Eggers², Reza Kazemi², Lukas Glos², Emily G. Ewing², Calvin B. Bannister², Keith Humphreys^{7,8}, Douglas C. Throckmorton², Mohammad S. Jalali^{1,3*}

¹Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139

²Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 20993

³MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA 02115

⁴Harvard Business School, Boston, MA 02163

⁵Homer Consulting, Barrytown, NY 12507

⁶Systems Science Program, Portland State University, Portland, OR 97201

⁷Veterans Affairs Health Services Research Center, Palo Alto, CA 94304

⁸Department of Psychiatry and Behavioral Sciences, Stanford University Medical Center, Stanford, CA 94305

*Corresponding author: Mohammad S. Jalali

Email: MSJALALI@mgh.harvard.edu

This PDF file includes:

- S1) Glossary
- S2) Full Model Structure
- S3) Data Sources
- S4) Model Estimation
- S5) Full Results
- S6) Sensitivity Analyses
- Figs. S1 to S21
- Tables S1 to S17
- SI References

Other supplementary materials for this manuscript include the following:

Online repository at <https://github.com/FDA/SOURCE> (most up-to-date active version) and <https://zenodo.org/record/6544836> (archived version for this publication)

Table of Contents

S1)	Glossary.....	6
S1.a)	Definitions of key terms.....	6
S1.b)	List of acronyms.....	7
S2)	Full Model Structure.....	8
S2.a)	Overview of structure.....	8
S2.b)	Model states & transitions.....	8
S2.c)	Major feedback effects.....	12
S2.c.i)	Social influence.....	13
S2.c.ii)	Perceived risk.....	14
S2.c.iii)	Availability.....	15
S2.c.iv)	Inclusion & exclusion of specific feedback effects.....	16
S2.d)	Additional model sectors.....	17
S2.d.i)	Opioid supply & availability.....	17
S2.d.ii)	Treatment.....	19
S2.d.iii)	Overdoses, naloxone, and synthetics.....	25
S3)	Data Sources.....	28
S3.a)	Main drug use states & transitions.....	28
S3.a.i)	Prescription opioid misuse.....	29
S3.a.ii)	Prescription opioid use disorder.....	29
S3.a.iii)	Non-disordered heroin use.....	30
S3.a.iv)	Heroin use disorder.....	30
S3.a.v)	Initiating prescription opioid misuse.....	32
S3.a.vi)	Initiating heroin use.....	32
S3.a.vii)	Heroin use adjustments.....	32
S3.a.viii)	Total population.....	33
S3.b)	Opioid prescribing, supply, and price data.....	33
S3.b.i)	Prescription opioid supply.....	33
S3.b.ii)	Prescription opioid demand.....	36
S3.b.iii)	Heroin price.....	36
S3.c)	Treatment & remission.....	36
S3.c.i)	Treatment receipt.....	36
S3.c.ii)	Treatment-seeking and barriers.....	37
S3.c.iii)	Treatment capacity.....	39
S3.c.iv)	Treatment duration, outcomes, and effects.....	40
S3.c.v)	Remission.....	41
S3.d)	Overdoses, naloxone, and synthetics.....	42
S3.d.i)	Overdose mortality data.....	42
S3.d.ii)	Synthetic death allocation structure.....	43
S3.d.iii)	Fentanyl penetration.....	43
S3.d.iv)	Naloxone distribution.....	44
S3.d.v)	Intervention Probabilities.....	45
S3.e)	Literature sources for parameter estimates.....	45
S3.f)	Expert consultation process.....	46
S4)	Model Estimation.....	47
S4.a)	Overview.....	47
S4.a.i)	Likelihood function for historical data.....	47

S4.a.ii)	Penalty terms for key point observations.....	48
S4.b)	Data used in estimation	48
S4.c)	Iterative estimation procedure	50
S4.d)	Quantifying uncertainties	52
S4.d.i)	Credible intervals for parameter estimates.....	52
S4.d.ii)	Estimating measurement error.....	52
S4.e)	Synthetic data validation	53
S4.e.i)	Data generation and estimation	53
S4.e.ii)	Synthetic data estimation results	53
S4.f)	Out-of-sample validation	54
S4.f.i)	Estimation & assessment process.....	55
S4.f.ii)	Out-of-sample validation results	55
S5)	Full Results	56
S5.a)	Full parameter estimates	56
S5.b)	Fit to historical data	57
S5.c)	Base case projections.....	59
S5.d)	Full synthetic data estimation results.....	63
S5.e)	Full out-of-sample validation results	66
S6)	Sensitivity Analyses.....	68
S6.a)	Sensitivity of outcomes to parametric assumptions.....	68
S6.a.i)	Parametric sensitivity results.....	69
S6.b)	Feedback loop knockout analyses	71
S6.b.i)	Loop knockout results.....	71
S6.c)	Sensitivity of projections to exogenous input assumptions	76
S6.d)	Sensitivity of parameter estimates to inclusion of 2020 data	78
S7)	Model Equations	79
S8)	References	81

List of Figures

Figure S1. Overview of model use states (stocks) and transitions (flows).	9
Figure S2. Overview of key feedback effects in model.....	12
Figure S3. Use state populations driving each social influence effect.	14
Figure S4. Feedback effects actively or potentially influencing each transition.	16
Figure S5. Treatment engagement as a dual continuum of care.	20
Figure S6. Functional relationship between buprenorphine-waivered providers and marginal effective capacity added per new waived provider.	23
Figure S7. Functional relationship between duration of treatment and treatment success fraction.	24
Figure S8. Functional relationship between naloxone kits distributed and probability of naloxone administration.....	26
Figure S9. Prescription opioid / heroin use state matrix with corresponding NSDUH data variables or model states.....	28
Figure S10. Observed patients per provider against number of waived providers in historical data	39
Figure S11: Allocation of overdose mortality by MECE categories to model data streams.....	42
Figure S12. Overview of iterative estimation procedure	50
Figure S13. Theoretical vs. actual percentage of parameters contained within different credible intervals.	54
Figure S14. Comparison of simulated model output (blue) to historical data (grey).....	58
Figure S15. Base case ('exogenous trends continue') projections for all time-series data used in estimation.	61
Figure S16. 'Optimistic' case projections for all time-series data used in estimation.....	62
Figure S17. 'Pessimistic' case projections for all time-series data used in estimation.....	63
Figure S18. Estimated values & 95% credible intervals for all synthetic data parameters compared with 'true' values	66
Figure S19. Comparison of model projections (blue) against historical data (grey), showing predictive accuracy for out-of-sample validation test. Bands are 95% CrIs for estimated underlying values (estimation portion, before 2012) and for projected reported data (after 2012); projected reported values account for measurement noise, which is present in the actual holdout data.	67
Figure S20. Model fit to data for loop knockout analyses, for all time-series data used in estimation.....	75
Figure S21. Constant-assumptions case projections for all time-series data used in estimation.	78

List of Tables

Table S1. Main states, transitions, and feedback coefficients	11
Table S2. DSM-5 criteria for substance use disorder, compared to DSM-IV substance abuse & substance dependence criteria.....	30
Table S3. MME conversion factors for different opioid substances	34
Table S4. FDA-approved abuse-deterrent formulation opioids currently marketed in the U.S.*	35
Table S5. Rates of treatment-seeking by use disorder and MOUD type, relative to total Tx-seeking rate for OUD (r_{UTR}).....	37
Table S6. Barriers to treatment engagement queried in NSDUH (with variable codes)	38
Table S7. Literature sources for parameter estimates.....	46
Table S8. Panel of time-series data used in model estimation	49
Table S9. Point data used as priors y_{kt} in model estimation	50
Table S10. Complete list of estimated parameter values & credible intervals	56
Table S11. Goodness of fit statistics for each estimated time series	59
Table S12. Simulated values for key point observations.....	59
Table S13. Exogenous inputs with alternative base case assumptions for projections	60
Table S14. Sensitivity of key outcomes to $\pm 10\%$ changes in parametric assumptions.	70
Table S15. Sensitivity of key outcomes to loop knockout analyses.	73
Table S16. Sensitivity of key projected outcomes to alternative base case assumptions	77
Table S17. Sensitivity of key outcomes to exclusion of 2020 data.....	80

S1) Glossary

S1.a) Definitions of key terms

Prescription opioids: Prescription (Rx) opioids are analgesic medications, used primarily to treat pain. Examples include natural opiates such as morphine or codeine; semi-synthetic opioids such as hydrocodone and oxycodone; and synthetic opioids such as tramadol and licit fentanyl (see also **fentanyl** below). They most commonly come in pill form, though other forms (e.g., liquid, film, etc.) exist as well. Prescription opioids are pharmaceutical products, though illicitly manufactured counterfeit prescription pills, often containing fentanyl, are a growing concern (see S2.d.i.(2)). We use the term ‘prescription opioids’ to refer to any pharmaceutically produced opioid analgesic, regardless of how it is obtained or used (e.g., whether prescribed by a medical provider or diverted; whether used to treat pain as prescribed or for other purposes).

Heroin: Heroin is an illicit semi-synthetic opioid that comes in several forms (e.g., black tar, brown or powder). It is consumed in several ways, including oral intake, snorting, smoking, and injection. As an illicit drug, the production, distribution, and sale of heroin is illegal. Heroin is often contaminated with various adulterants, and increasingly with **fentanyl** (see below).

Fentanyl: Fentanyl is a highly potent synthetic opioid with many analogues (e.g., carfentanil, sufentanil, etc.). While licit, pharmaceutically produced prescription fentanyls exist, they are relatively uncommon; the majority of fentanyl in circulation now is illicitly manufactured fentanyl (IMF). IMF is increasingly common in the supply of **heroin** and other illicit drugs (see S2.d.iii.(3)). Other, non-fentanyl synthetic opioids exist as well, though they are generally less potent and far less common in the illicitly-manufactured fentanyl supply, which includes both basic fentanyl and numerous analogues. In this model, we do not distinguish between them, and use the terms ‘synthetic[s]’ and ‘fentanyl’ interchangeably to refer to illicitly manufactured fentanyl and its analogues, unless otherwise specified. We specifically use the terms ‘prescription synthetics’ or ‘prescription fentanyl’ to refer to the licit form (see **prescription opioids**).

Misuse: Prescription opioid misuse includes any use of Rx opioids prescribed for someone else, or use of Rx opioids solely ‘for the feeling [they] caused’ (see S3.a.i)). As an umbrella term, ‘misuse’ can also include ‘low-intensity’ use of heroin that does not rise to the level of **use disorder** (see below), which we also term ‘non-disordered heroin use’ (NDHU; see S3.a.iii)).

Use disorder: Substance use disorder is a clinically-diagnosable psychiatric disorder defined in the DSM-5 (see S3.a.ii)). Use disorder of varying degrees of severity is defined by endorsement of an increasing number of criteria identifying problems associated with drug use. Substance use disorder is associated with use of a particular substance; we distinguish between ‘Rx opioid use disorder’ and ‘heroin use disorder’ (see S3.a.iv)).

Remission: Remission is the reduction or disappearance of symptoms of **use disorder**. An individual who formerly qualified as having use disorder and now no longer meets the criteria for use disorder is in remission. While the term ‘recovery’ is used more generally to refer to the process of going from use disorder to a normal state of functioning and quality of life (see S3.c.v)), we focus on ‘remission’ as defined relative to use disorder. Note that remission does not necessarily entail complete abstinence from substance use.

Medication for Opioid Use Disorder (MOUD): Medication[s] for opioid use disorder [MOUD] refers to one or more of a set of three FDA-approved medications used to treat OUD – buprenorphine, methadone, and Vivitrol®. Treatment with MOUD is sometimes referred to as medication-assisted treatment (MAT) or opioid agonist therapy (OAT). There are many forms of treatment for use disorder, e.g., psychosocial therapy, community support groups, 12-step programs, etc. in addition to treatment with MOUD. However, our model explicitly represents MOUD but not other forms of treatment (see S2.b)); we therefore sometimes use ‘MOUD’ and ‘treatment’ interchangeably in the context of the model to refer to treatment involving MOUD.

S1.b) List of acronyms

ADF	Abuse-deterrent formulation
Bup / Bupe	Buprenorphine
CDC	Centers for Disease Control and Prevention
CMS	Centers for Medicare & Medicaid Services
DEA	Drug Enforcement Administration
DSM (DSM-IV / DSM-5)	Diagnostic and Statistical Manual of Mental Disorders (Fourth / Fifth edition)
EMS	Emergency medical services
ETC	Exogenous trends continue
FDA	Food and Drug Administration
H	Heroin
HHS	Department of Health and Human Services
HUD	Heroin use disorder
ICD (ICD-9 / ICD-10)	International Classification of Diseases (9 th / 10 th edition)
IMF	Illicitly manufactured fentanyl
MME	Milligrams morphine equivalent
MMT	Methadone maintenance therapy
MOUD	Medications for opioid use disorder
NASEM	National Academies of Sciences, Engineering, and Medicine
NCHS	National Center for Health Statistics
NDHU	Non-disordered heroin use
NESARC	National Epidemiologic Survey on Alcohol and Related Conditions
NFLIS	National Forensic Laboratory Information System
NSDUH	National Survey on Drug Use and Health
NSDUH RDAS	NSDUH Restricted-use Data Analysis System
N-SSATS	National Survey of Substance Abuse Treatment Services
NVSS	National Vital Statistics System
Nx	Naloxone
OD	Overdose
OSM	Opioid systems model
OUD	Opioid use disorder
Rx	Prescription / prescription opioid[s]
Rx OUD	Prescription opioid use disorder
SAMHSA	Substance Abuse and Mental Health Services Administration
STRIDE	System to Retrieve Information from Drug Evidence
SUD	Substance use disorder
TEDS	Treatment Episode Data Set
Tx	Treatment (for use disorder)
UNODC	United Nations Office on Drugs and Crime
Viv	Vivitrol® (naltrexone)

S2) Full Model Structure

S2.a) Overview of structure

SOURCE is a continuous-time differential equations model, developed using a systems approach that emphasises endogenous feedback processes within a broad model boundary that drive changes over time. The model simulates the movement of people through different states of opioid use, with endogenous influences on initiation and transition rates, as well as more detailed representations of prescribing, treatment, and overdose-related processes. Broadly, we distinguish people by severity of opioid use (misuse vs. use disorder), as well as by substances used (prescription vs. illicit opioids) (see S3.a)). The model is parametrised to represent the opioid-using population in the U.S. at a national level. Here we present key equations and structures in each of its sectors, with a complete listing of model equations in S7). Data sources for each sector are detailed in S3).

The model was developed and implemented using Vensim™ simulation software; all model files are available in the online repository at <https://github.com/FDA/SOURCE>.

S2.b) Model states & transitions

The opioid system includes people in various stages of use of both prescription opioids (*Rx*) and illicit opioids like heroin (*H*). For all disorder and remission definitions, we use the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) criteria (1). **Figure S1** provides an overview of the key population groups and the transitions among them (for more detailed definitions of these states and corresponding data sources, see S3.a)).

People enter the opioid system by either initiating prescription opioid misuse – with their own prescription (*initiating Rx misuse own Rx*, r_{MI}), or with others' (*initiating Rx misuse diverted*, r_{MD}) – or by initiating heroin use without prior Rx opioid misuse (*initiating heroin no Rx*, r_{ND}).¹ Definitions of opioid misuse vary; we follow the 2002-2014 NSDUH definition, to include *any* use of someone else's opioid prescription, *or* use of Rx opioids solely 'for the feeling [they] caused' (2).

People who initiate Rx opioid misuse enter the stock of people with *Rx misuse* (*M*), while people who initiate heroin use without prior Rx misuse enter the stock of people with *non-disordered heroin use* (*N*). People misusing opioids can also initiate heroin (*initiating heroin with Rx misuse*, r_{MN}) and enter *N*. Once people transition from *M* to *N*, they are no longer distinguished from people who transitioned directly into *N* without first using Rx opioids. People in *M* and *N* can quit use in a given year, but also later resume use, with net flows (*net quitting Rx misuse*, r_{MQ} ; *net quitting NDHU*, r_{NQ}) reflecting the combined total of quits and resumptions of use (but not new initiations) at any given time.

¹ The growing presence of illicit fentanyl in the heroin supply complicates identification of 'heroin use', with heroin almost completely displaced by fentanyl in some parts of the country (206). To our knowledge, few if any people self-identify as users of fentanyl as distinct from heroin, and use behaviours are similar in any case. As such we use 'heroin' to refer more accurately to illicit opioids, typically not in pill form, that may contain illicitly manufactured fentanyl in addition to or instead of heroin; this includes any powder-form drug that users reasonably think contains heroin or fentanyl, as well as non-powder-form (e.g., black tar) heroin.

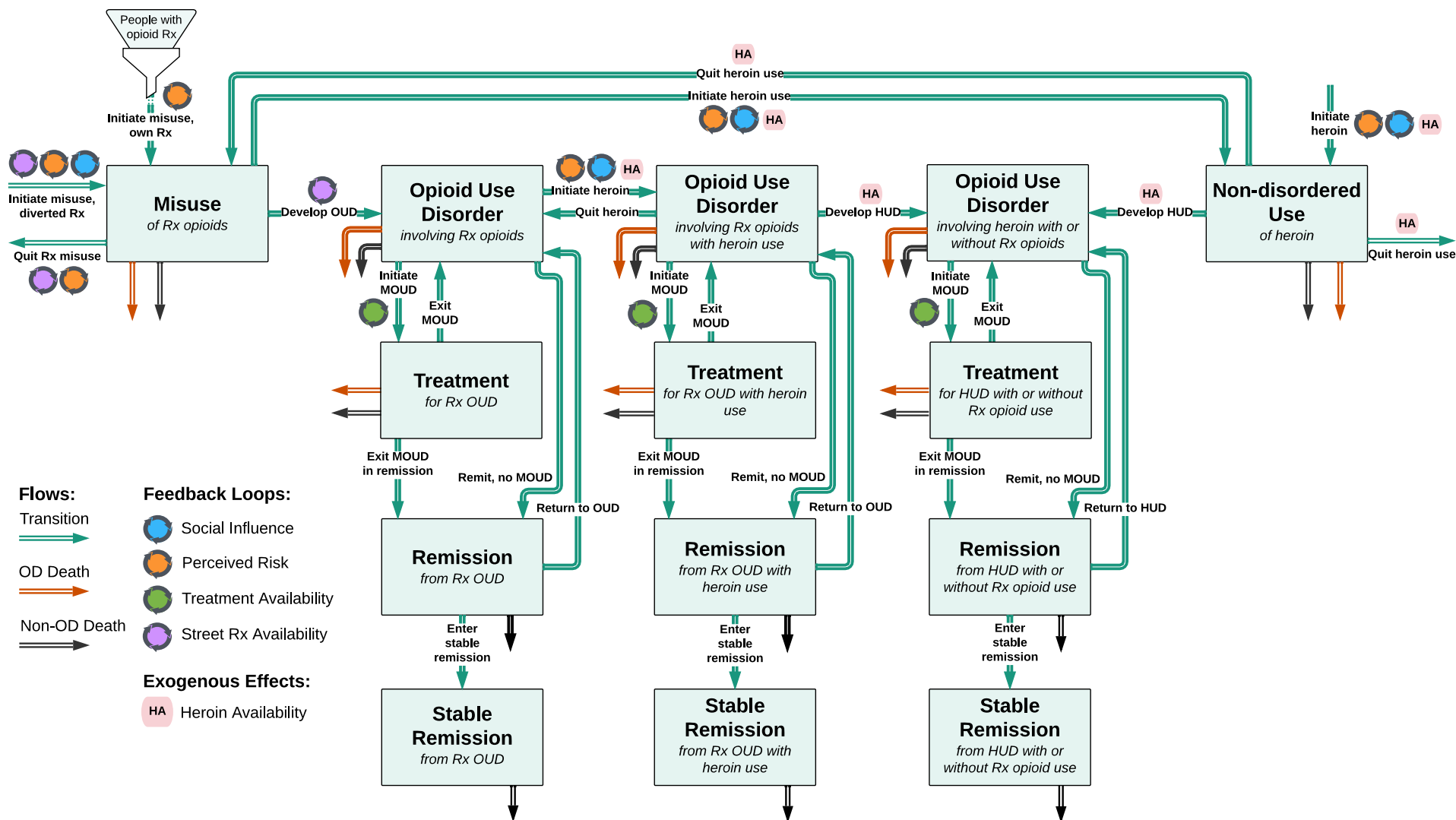


Figure S1. Overview of model use states (stocks) and transitions (flows). Treatment states are further separated by MOUD type.

From Rx misuse, M , people can develop opioid use disorder (OUD) (*developing Rx OUD*, r_{MU}), thereby entering a disordered state involving Rx opioids only (*Rx OUD no PY heroin*, U_R). From the non-disordered heroin use state, they can develop an OUD involving heroin (*developing HUD no Rx OUD*, r_{NU}). For clarity, we call this state *heroin use disorder* or HUD (U_H). We also distinguish a third use disorder state, *Rx OUD with PY heroin* (U_O), which encompasses people with Rx OUD who have also used heroin in the past year, but whose heroin use does not rise to the level of a use disorder. While relatively uncommon, this is an important transitional state, which we therefore represent explicitly. People enter this state from U_R by *initiating heroin with Rx OUD* (r_{UO}). Once in this state, people can also develop HUD (*developing HUD with Rx OUD*, r_{OH}).²

Once in the use disorder states (U), people enter remission (*... in remission*, $R_{(.)}$) through one of two pathways: via remission without use of medications for opioid use disorder (*remitting... no MOUD*, $r_{UR(.)}$), which could include psychosocial or behavioural treatment or no treatment at all; or through treatment with MOUD (*... in MOUD Tx*, $T_{(.)}$). Remission occurs after no longer meeting criteria for a DSM-5 disorder for at least one year. Once in remission, the probability of relapse (*relapsing...*, $r_{RU(.)}$) or remaining in remission is the same regardless of the pathway by which remission was achieved, with or without MOUD.

After some time in the remission states, people transition to a more durable state of stable remission (*... in stable remission*, $R_{S(.)}$), from which we assume they are no longer at risk of relapse. This transition (*stabilizing remission...*, $r_{RS(.)}$) takes place after several years (*time to stabilize remission*, τ_{RS}) in the base remission state (see S3.c.v)).

Treatment engagement can involve any of the three FDA-approved MOUDs: buprenorphine, methadone, and Vivitrol (subscripts B , M , V respectively). *Treatment engagement* flows ($r_{UT(.)}$) are limited by both demand and capacity for each of these medications separately, as explained further in S2.d.ii). Once in treatment, people can leave treatment before remitting, thereby returning to use disorder (*Tx exit with UD*, $r_{TU(.)}$), or leave in remission (*Tx exit in remission*, $r_{TR(.)}$). Throughout the use disorder-treatment-remission chain, stocks are separated by drugs of use (subscripts R , O , H) and by medication used (subscripts B , M , V) as appropriate.

Each stock in the model also has two additional outflows (one for remission states, $R_{(.)}$) – death from non-overdose causes (*nonOD death*, $n_{(.)}$), as well as opioid-caused *overdose death* ($o_{(.)}$) for all states except remission. Overdose death rates are significantly impacted by naloxone availability and fentanyl penetration into the heroin supply, as detailed in S2.d.iii).

The vast majority of transition rates or flows in the model are formulated as fractional annual hazard rates ($\rho_{(.)}$) multiplied by source populations, sometimes further multiplied by additional coefficients, e.g.:

$$r_{MN} = (\rho_{MN}M)S_{MN}P_{MN}A_{MN} \quad (2.1)$$

Where $S_{(.)}$, $P_{(.)}$, and $A_{(.)}$ are coefficients for various endogenously generated effects, elaborated on in S2.c). Different transition rates are subject to different effects and coefficients, while treatment entry/exit and overdose death flows are subject to additional influences as well. In most cases, the base rates (ρ) are estimated model parameters; in a few cases they are derived from extant literature.

² Note that the distinction between use disorder states is based on substance[s] of use and use behaviours, not the sources of those substances. See S3.a) for details.

For two of the three entry flows into the system (r_{MD} , r_{ND}), source populations are not explicitly represented in the model. Instead, we estimate an absolute base rate in place of a fractional hazard rate, which implicitly accounts for the source population size. For the third entry flow, misuse starting with one's own prescription (r_{MI}), we calculate the number of medical users of Rx opioids (*patients with current month opioid Rx*, m_c) as the source population-based on exogenous input data. The patient population is large and relatively static compared to the rest of the model, with very short average residence times, so it is not explicitly modelled as a stock. Details of this calculation are included in S3.b).

Table S1. Main states, transitions, and feedback coefficients

State variables		
M	Rx misuse no heroin	Prescription opioid misuse
N	Nondisordered heroin use	Non-disordered heroin use
U _R	Rx OUD no PY heroin not in MOUD Tx	Prescription opioid use disorder, no past-year heroin use
U _O	Rx OUD with PY heroin not in MOUD Tx	Prescription opioid use disorder, past-year heroin use
U _H	HUD not in MOUD Tx	Heroin use disorder
T _R	Rx OUD no heroin by MOUD	Prescription opioid use disorder, no past-year heroin use, in medication for opioid use disorder treatment
T _O	Rx OUD with heroin by MOUD	Prescription opioid use disorder with past-year heroin use, in medication for opioid use disorder treatment
T _H	HUD by MOUD	HUD in medication for opioid use disorder treatment
R _R	Rx OUD no heroin in remission	Remission from prescription opioid use disorder, no heroin use in the year prior to quitting
R _O	Rx OUD with heroin in remission	Remission from Rx OUD with heroin use in the year prior to quitting
R _H	HUD in remission	Remission from heroin use disorder
R _{SR}	Rx OUD no heroin in stable remission	> 5 years in remission from prescription opioid use disorder, no heroin use in the year prior to quitting
R _{SO}	Rx OUD with heroin in stable remission	> 5 years in remission from Rx OUD with heroin use in the year prior to quitting
R _{SH}	HUD in stable remission	> 5 years in remission from heroin use disorder
Transitions		
r_{MI}	Initiating Rx misuse own Rx	Initiating prescription opioid misuse with one's own prescription opioid
r_{MD}	Initiating Rx misuse diverted	Initiating prescription opioid misuse with someone else's prescription opioid
r_{ND}	Initiating heroin no Rx	Initiating non-disordered heroin use without having misused prescription opioids
r_{MN}	Initiating heroin with Rx misuse	Initiating non-disordered heroin use after having misused prescription opioids
r_{MQ}	Net quitting Rx misuse	Quitting prescription opioid misuse
r_{NQ}	Net quitting NDHU	Quitting non-disordered heroin use
r_{MU}	Developing Rx OUD	Developing opioid use disorder from prescription opioid use
r_{NU}	Developing HUD no Rx OUD	Developing opioid use disorder from heroin use without having had opioid use disorder from prescription opioid use
r_{UO}	Initiating heroin with Rx OUD	Initiating heroin use after having had opioid use disorder from prescription opioid use
r_{OH}	Developing HUD with Rx OUD	Developing opioid use disorder from heroin after having had opioid use disorder from prescription opioid use

$r_{UR(.)}$	Remitting... no MOUD	Remitting from (...) without medication-based treatment for opioid use disorder
$r_{RU(.)}$	Relapsing...,	Returning to opioid use disorder from remission from (...)
$r_{UT(.)}$	Treatment engagement	Engaging in medication-based treatment for opioid use disorder from (...)
$r_{TU(.)}$	Tx exit with UD	Exiting medication-based treatment for opioid use disorder from (...) with opioid use disorder from (...)
$r_{TR(.)}$	Tx exit in remission	Exiting medication-based treatment for opioid use disorder for (...) in remission from (...)
$n_{(.)}$	<i>NonOD death</i>	Dying from causes besides opioid-involved overdose from (...)
$o_{(.)}$	<i>Overdose death</i>	Dying from an opioid-involved overdose from (...)
Feedback effects		
$S_{(.)}$	<i>Social influence coefficient</i>	Effect of social influence processes on (...)
$P_{(.)}$	<i>Perceived risk coefficient</i>	Effect of responses to perceived risk on (...)
$A_{(.)}$	<i>Rx / Heroin / Rx vs. H availability coefficient</i>	Effect of drug availability or comparative availability on (...)

S2.c) Major feedback effects

The model contains three main sets of endogenous influences (i.e., feedback loops or effects) on transition rates ($r_{i,j}$) between use states, shown in **Figure S2**:

- 1) Social influence reinforcing feedbacks, whereby existing users increase initiation and people with UD accelerate disorder development among existing users;
- 2) Risk perception balancing feedbacks, whereby opioid overdoses, especially overdose mortality, discourage initiation;
- 3) Availability balancing feedbacks, whereby the availability of Rx opioids (sometimes compared to heroin) fluctuates with the balance of supply and demand, influencing initiation, development of use disorder, transitions between Rx opioid and heroin use, and potentially quitting.

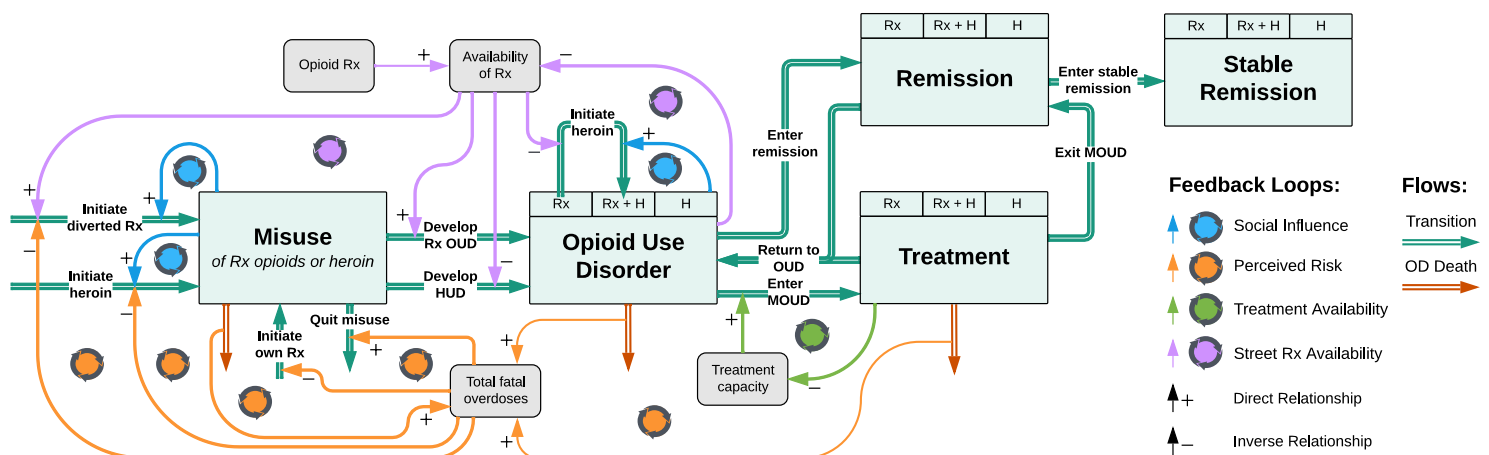


Figure S2. Overview of key feedback effects in model

These feedback effects are all formulated with the same basic structure:

$$C_{(.)} = (D_{C(.)})^{\varepsilon_{(.)}} \quad (2.2)$$

Where $C_{(.)} \in \{S_{(.)}, P_{(.)}, A_{(.)}\}$ is the *social influence coefficient*, *perceived risk coefficient*, or *Rx availability / H availability / Rx vs. H availability coefficient* for a given transition rate respectively; $D_{C(.)}$ is the relevant driver of the effect (*relative social influence*, *relative perceived risk*, *relative availability*); and $\varepsilon_{(.)} \in \{\psi_{(.)}, \pi_{(.)}, \alpha_{(.)}\}$ is the *social influence strength*, *perceived risk strength*, *availability strength* for that particular transition rate. These effect strengths are model parameters, estimated through the model estimation process (see S4)).

Specifics on the drivers of each effect are in the following sections; in all cases, the driver is a time-varying quantity normalised by its initial value. Normalising allows coefficients on transition rates to vary with changes in their drivers without needing to tease apart baseline transition rates from the endogenous effects present at the start of the simulation time period.

In addition to these three main sets of feedbacks, treatment capacity limitations create a fourth, balancing feedback process, whereby as new patients enter treatment, the limited number of available treatment spots is filled, reducing or preventing further treatment engagement until existing patients leave (see S2.d.ii.(2)).

S2.c.i) Social influence

Drug use behaviour has an element of social contagion (3–7). As more people use a substance, its use becomes increasingly normalised, and relevant knowledge about its use (e.g., methods of administration, sources of supply, etc.) becomes more widespread and accessible (4, 7). Access to the substance in social networks grows as people seek the substance or become suppliers to others (especially in the case of prescription opioids). Collectively, these processes increase initiation of drug use, creating a self-reinforcing growth process. (These processes can also work in reverse as use declines.) Similarly, social space-driven ratcheting effects can potentially drive increasingly heavy drug use (8, 9), which we operationalise as social influence on the UD development process (but see S2.c.iv)).

We operationalize social influence separately for Rx misuse (r_{MI} , r_{MD}) and heroin initiation (r_{ND} , r_{MN} , r_{UO}) flows, as well as for initiation vs. development of use disorder (r_{MU} , r_{NU} , r_{OH}). The *relative social influence* ($D_{S(.)}$) for a given transition depends on the fraction of the total population (see S3.a.viii)) engaging in the relevant drug use behaviours. Essentially, 1) only users of a given substance class (Rx vs. heroin) exert social influence on initiation or development of use disorder for that substance, and 2) heavier users exert influence on lighter users, but not vice-versa, such that people with use disorder affect initiation rates, but those without use disorder do not affect use disorder development rates (see **Figure S3** for details).

Note that while network effects on the accessibility of drugs in social networks are captured in this social influence process, the aggregate effect of a changing user base on the demand-supply balance is represented separately in the availability effects detailed below.

Flow Symbol	Flow Name	Population exerting social influence				
		Rx misuse no heroin (M)	Rx OUD no PY heroin not in MOUD Tx (U_H)	Rx OUD with PY heroin not in MOUD Tx (U_O)	HUD not in MOUD Tx (U_H)	Nondisordered heroin use (N)
r_{MD}	Initiating Rx misuse diverted					
r_{ND}	Initiating heroin no Rx					
r_{MN}	Initiating heroin with Rx misuse					
r_{MU}	Developing Rx OUD					
r_{NU}	Developing HUD no Rx OUD					
r_{UO}	Initiating heroin with Rx OUD					
r_{OH}	Developing HUD with Rx OUD					

Figure S3. Use state populations driving each social influence effect. Initiating Rx misuse from diverted opioids is influenced by the fraction of people in the non-disordered heroin use state who also misuse Rx opioids.

S2.c.ii) Perceived risk

Perceived risk coefficients $P_{(.)}$ reflect the deterrent effect that adverse outcomes like death can have on drug use behaviour. As overdoses and especially overdose mortality become more common, the perceived risk associated with a drug increases, dissuading potential initiates (reducing r_{MI} , r_{MD} , r_{ND} , r_{MN} , r_{UO}) and possibly encouraging current misusers (but not people with a disorder) to quit use (r_{MQ} , r_{NQ}), creating a balancing feedback process (10–12).

The perceived risk associated with use of a drug (D_{PR} , D_{PH}) adjusts with some lag to an underlying *indicated perceived risk* (D_{PR}^* , D_{PH}^*). The lag is asymmetric, i.e., the *perceived risk increase time* (τ_{PI}) is significantly shorter than the *perceived risk decrease time* (τ_{PD}), reflecting that deaths, overdoses, etc. tend to get more attention than the lack of them, and a dangerous reputation for a drug fades slowly (10, 11). The indicated perceived risk is operationalized as a weighted sum of the fatal and nonfatal overdoses associated with that drug, with a lower relative weight (*perceived risk weight NFOD*, w_η) given to non-fatal overdoses in users' or potential users' perceptions of risk:

$$\frac{dD_P}{dt} = \frac{D_P^* - D_P}{\tau_P} \quad (2.3)$$

$$D_P = \sum_{(.)} o_{(.)} \left(1 + w_\eta \frac{\eta_{(.)}}{\omega_{(.)}} \right), \quad (.) \in \{R, H\} \quad (2.4)$$

Where $\eta_{(.)}/\omega_{(.)}$ is the ratio of nonfatal to fatal overdoses (see S2.d.iii.(4)). Nonfatal overdoses are far more common and receive far less attention (especially for people not already using drugs) than fatal overdoses, so we assume a value of 0.1 for w_η , i.e., nonfatal overdoses carry 10% the risk perception impact of fatal overdoses.

Note that we operationalise perceived risk based on overdoses, i.e., adverse health outcomes, and do not incorporate perceived risk of arrest, incarceration, or other legal consequences. Rational choice models of illegal behaviour posit that legal consequences could raise the expected costs of drug use and thereby exert a deterrent effect (13, 14). However, there is little evidence that people who use or initiate drugs behave as rational actors in this way (14–17). Absent such evidence, we have excluded legal risks from the risk perception feedback effect.

S2.c.iii) Availability

Availability coefficients A_{\odot} represent the effects of market forces and drug supply on initiation, use disorder development, and quit rates. The availability of Rx opioids (*Rx availability for misuse*, D_{ARM}) affects initiation of Rx misuse and development of Rx OUD (r_{MD} , r_{MU} , r_{MQ}).

Rx availability is in part a function of demand for Rx opioids, which in turn depends on the number of users. It thus exerts a balancing effect whereby more people using reduces the relative availability, in turn reducing initiation. Numerous other factors also influence availability, as detailed in S2.d.i).

Similarly, the availability of heroin (*heroin availability index*, D_{AH}) can exert an effect on heroin initiation and use disorder development flows (r_{ND} , r_{NU} , r_{NQ}), with greater availability facilitating initiation and UD development and discouraging quitting. Note, however, that we model heroin availability exogenously (see S2.d.i.(3)), so this is not, strictly speaking, a feedback process.

In addition to the separate availabilities of Rx opioids and heroin, we also consider their comparative availability, which affects transitions between Rx and heroin use. For purposes of this comparison, we use separate Rx availability constructs for prescription opioid misuse vs. use disorder (D_{ARM} vs. *Rx availability for UD*, D_{ARU}), as detailed in S2.d.i.(2) below. The ratio of the respective Rx availability construct to heroin availability yields the *Rx vs heroin availability index misuse* (D_{ACM}), which drives heroin initiation from Rx opioid misuse (r_{MN}), or the *Rx vs heroin availability index UD* (D_{ACU}), which drives initiation or escalation of heroin use with Rx OUD (r_{UO} , r_{OH}).

S2.c.iii.(1) ADF effects on heroin initiation

In addition to availability effects, we allow for one additional supply-related effect on transitions – an effect of abuse-deterrent formulations (ADFs) on heroin initiation with Rx OUD (r_{UO}). Like heroin availability effects, this is not strictly speaking a feedback process, but operates in a similar way, driven by the *ADF fraction of Rx street supply* (F^{AS}) (see S2.d.i.(2)).

ADF prescription opioids are specially formulated to impede physical or chemical modification (e.g., crushing or dissolving), which makes them less amenable to non-oral routes of administration (e.g., snorting or injecting) (18). In principle, the intended effect of ADFs is to deter escalation from oral to non-oral misuse of prescription opioids. We do not explicitly distinguish between routes of administration in this model, and therefore cannot represent this effect directly. However, non-oral misuse of opioids is a marker of OUD severity and a significant predictor of heroin initiation (19, 20). We therefore approximate the potential effect of ADFs on reducing non-oral misuse as an effect on the subsequent transition to heroin use instead.

S2.c.iv) Inclusion & exclusion of specific feedback effects

The feedback processes explained above are all plausible influences on opioid use transitions, with some evidence for their effects. However, the magnitude of each effect and its impact on, e.g., initiation rates is difficult to discern with precision from available evidence. For instance, surveys of attitudes toward drug use among young people indicate an increase in the perceived risk associated with Rx opioids and heroin over the last decade³, but do not associate those changing attitudes with changing likelihoods of initiating drug use. We therefore need to ascertain the impact of each process from the aggregate data, through model estimation.

In order to allow the potential impact of each feedback to emerge from the data, we include all the aforementioned plausible feedbacks in the model structure during the estimation process. Some of the resultant estimated effect strengths show no significant effect for a given feedback on a given rate ($\varepsilon_{(.)} \sim 0$); those specific feedbacks are thus inactive in the final model.

Flow Symbol	Flow Name	Social influence	Perceived risk	Availability effects		
				Rx availability	Heroin availability	Rx vs. heroin availability
r_{MI}	Initiating Rx misuse own Rx					
r_{MD}	Initiating Rx misuse diverted					
r_{ND}	Initiating heroin no Rx					
r_{MN}	Initiating heroin with Rx misuse					
r_{MO}	Net quitting Rx misuse					
r_{NO}	Net quitting NDHU					
r_{MU}	Developing Rx OUD					
r_{NU}	Developing HUD no Rx OUD					
r_{UO}	Initiating heroin with Rx OUD					
r_{OH}	Developing HUD with Rx OUD					

Figure S4. Feedback effects actively or potentially influencing each transition. Initiating heroin with Rx OUD (r_{UO}) also includes a potential effect from ADFs (see S2.c.iii.(1))

In some cases, the lack of effect is likely due to under-determination. For instance, the effect of perceived risk on initiating and quitting heroin use (r_{ND} and r_{NQ}) is similar, and given the absence of any reliable data

³ Specifically, among Monitoring The Future respondents aged 18-30, the fraction perceiving ‘great risk’ of taking narcotics other than heroin just once or twice has risen from approximately 40% in 2011 (when the question was first asked) to 46% in 2018 (207). The fraction reporting the same for trying heroin once or twice has risen from 60% in 1999 to 66% in 2018. In NSDUH, among those with an Rx OUD who had not yet used heroin, the fraction perceiving “great risk” in using heroin once or twice rose from 70% in 2011 to 81% in 2018 (208).

on quit rates, cannot be distinguished. Additional data would allow re-estimation and potentially re-inclusion of these effect strengths. **Figure S4** summarises which feedback effects were allowed to potentially operate on which transitions in the estimation process, and in the final model; see also S6.b).

S2.d) Additional model sectors

S2.d.i) Opioid supply & availability

S2.d.i.(1) Prescribing and supply

Opioid prescribing practices influence both the number of medical users of opioids (m_C) who may initiate opioid misuse (r_{MI}), and the availability of Rx opioids.

The number of medical users of Rx opioids (*patients with current month opioid Rx*, m_C) is very large relative to other populations in the model, and their average ‘residence time’ fairly short. As such, the population of medical users is close to stable at any given time. We therefore represent them not as an explicit state variable, but with an analytic approximation:

$$m_C = m_P m_N m_D \quad (2.5)$$

Where m_P is the total number of *patients receiving opioid prescription annual*, m_N is the number of *prescriptions per person*, and m_D is the *average days per prescription*, as detailed in S3.b.i). The number of medical users at any given time is thus in effect the product of the rate of people receiving prescriptions and their average duration of medical use, per Little’s Law (21). Note that unlike most actual stocks in the model, m_C does not represent medical use *within the past year*, but rather current-month ongoing use. As such, the transition rates reflecting hazard of misuse initiation from prior medical use (ρ_{MI}) or overdose death for medical users (o_{mC}) should be interpreted as hazard rates per person-year of ongoing medical use of prescription opioids.

The *Rx supply* (q_S) represents the total supply that could be made available for potential misuse and can be thought of as ‘excess’ pills not used as prescribed within the time period of the prescription, which therefore present potential opportunities for misuse. Supply is fundamentally a function of total amount of prescription opioid medications dispensed each year, but is potentially influenced by more granular prescribing practices. We distinguish several aspects of prescribing that contribute to total amount prescribed, analogous to the Kaya identity (22) – in its basic form, total supply in morphine milligrams equivalent (MME) is the product of patients receiving prescriptions each year (m_P), prescriptions per patient (m_N), days per prescription (m_D), and MME dosage per day (m_M):

$$q_S = m_P \times m_N \times m_D \times m_M \quad (2.6)$$

These different aspects of prescribing patterns do not necessarily have equal weight in determining the effective supply of Rx opioids, as usage and consumption patterns differ. Simply put, giving twice as many people half as many opioids each vs. giving half as many people twice as many prescriptions each vs. giving the same number of people half the prescriptions of twice the dosage, and so on, will not necessarily have the same effect on supply. To allow for this possibility, we operationalise supply with a number of *sensitivity of Rx supply* exponents ($s_{s(C)}$), representing the relative contribution of each factor to overall supply. Specifically:

$$q_S = m_P^{s_{sp}} \times m_N^{s_{sn}} \times m_D^{s_{sd}} \times m_M^{s_{sm}} \quad (2.7)$$

Where each factor $m_{(.)}$ is normalised to its initial value, and the sensitivity exponents $s_{s(.)}$ are normalised to have a mean of 1. In the absence of more specific evidence, we assume a baseline value of 1 for each exponent, giving equal importance to number of patients, number of prescriptions, duration of prescriptions, and daily opioid dosage prescribed, though the relative contributions of each factor could be adjusted to test different possibilities. In addition, other aspects of prescribing such as the number of pills (units) could potentially be incorporated into an expanded formulation for supply.

S2.d.i.(2) Availability and street supply

The availability of Rx opioids for potential misuse (*Rx availability for misuse*, D_{ARM}) is driven by the ratio of *Rx supply* to *Rx demand for misuse* (q_D):

$$D_{ARM} = \frac{q_S + w_C q_{SC}}{q_D} \quad (2.8)$$

$$q_D = \sum_S S_{(.)} q_{DS(.)}, \quad S \in \{M, N, U, T\} \quad (2.9)$$

Where the supply side is the sum of Rx supply (q_S) and *counterfeit supply* (q_{SC}), downweighted by some *counterfeit supply weight* (w_C). The presence of counterfeit Rx opioids in the street supply is a growing concern (23–29), but there are no estimates presently available of their actual prevalence. As such, we allow for the possibility of their contributing to supply, potentially downweighted to reflect lower desirability, but set their quantity to 0. Rx demand (q_D) depends on the sizes of the populations in each drug use state and the expected average demand for individuals in that state (see S3.b.ii)).

The *Rx availability for UD* (D_{ARU}) likewise depends on Rx supply, potential counterfeit supply, and demand, as well as an additional *Rx street supply disruption* factor (Z):

$$D_{ARU} = \frac{q_S + w_C q_{SC}}{q_D} (1 - Z) = D_{ARM} (1 - Z) \quad (2.10)$$

Rx street supply disruption (Z) is a state variable reflecting short-term perturbations, beyond the longer-term dynamics of supply and demand, which affect the street market for Rx opioids:

$$\frac{dZ}{dt} = q_Z - \frac{Z}{\tau_Z} \quad (2.11)$$

The degree of disruption increases as *Rx street supply shocks* (q_Z) occur. We include a single such shock – the 2010 withdrawal of the crushable form of OxyContin from production. OxyContin was by far the single most widespread formulation in the prescription opioid street supply at the time (see S3.b.i.(5)), and although it was replaced with an abuse-deterrent formulation, the withdrawal of the non-ADF form nonetheless represented a substantial disruption of available supply, as the crush-resistant ADF form is not a perfect substitute. Disruptions fade as suppliers find new sources and consumers adjust their consumption preferences to available alternatives; this is a gradual process, taking *time to readjust Rx street supply* (τ_Z).

We separate Rx availability for people with misuse vs. use disorder (D_{ARM} vs. D_{ARU}) in order to allow these street supply disruptions to affect the latter but not the former. People with OUD consume far more opioids than those only misusing; they are much more likely to obtain at least some of their drugs from the ‘street’ or black market, including purchasing drugs through monetary or equivalent transactions (30,

31); and they are more likely to have specific preferences for higher-dosage units or pills they can modify for non-oral routes of administration (e.g., crushing or dissolving) (32). As such, they are more vulnerable or sensitive to potential disruptions in prescription opioid availability, particularly as compared to the availability of alternatives like heroin.

We calculate the *ADF fraction of Rx street supply* (F^{AS}) as a function of the *ADF fraction of prescribed Rx opioids* (F^{AR}):

$$F^{AS} = (F^{AR})^{s_{af}} \quad (2.12)$$

Where s_{af} is the *ADF substitutability factor*, representing the ability of the street supply to preferentially take up or avoid ADFs, shifting the composition of the street supply to include disproportionately high or low amounts of ADFs compared to what is prescribed (F^{AR}). While we allow for this possibility of differential uptake, in the absence of evidence indicating a strong skew one way or the other, we set $s_{af}=1$ by default, resulting in ADFs being as prevalent in the street supply as in the prescribed supply. We treat prescribed ADF supply (F^{AR}) as exogenous (see S3.b.i.(4)).

S2.d.i.(3) Heroin availability

As described in S2.c.iii), heroin availability can influence heroin initiation or UD development. In reality, heroin availability depends not only on street price but also features such as convenience, reliability, purity, and safety of obtaining supply (33). However, to our knowledge, there are no reliable data on availability or a suitable proxy thereof, besides price. We therefore operationalise heroin availability as simply the inverse of normalised heroin price, as calculated in S3.b.iii).

There is some evidence that heroin supply chains benefit from learning or improving returns to scale (34, 35), as producers, traffickers, distributors and dealers improve the efficiency of their practices or overwhelm law enforcement efforts. These learning effects may be partly responsible for the decline in heroin prices particularly from the mid-2000s onward (36, 37). However, the dynamics of the heroin supply chain and market are outside the scope of this model. As such, we do not represent these dynamics explicitly, instead treating heroin price as exogenous.

S2.d.ii) Treatment

S2.d.ii.(1) Treatment seeking, demand, and engagement

The process by which people receive addiction treatment can be thought of as a continuum of care (**Figure S5**), with some portion of patients lost to care at each step of the continuum upstream of actual *treatment engagement* ($r_{UT(i)}$). We represent this continuum with multiple variables, replicated as appropriate for each use disorder and/or MOUD type.

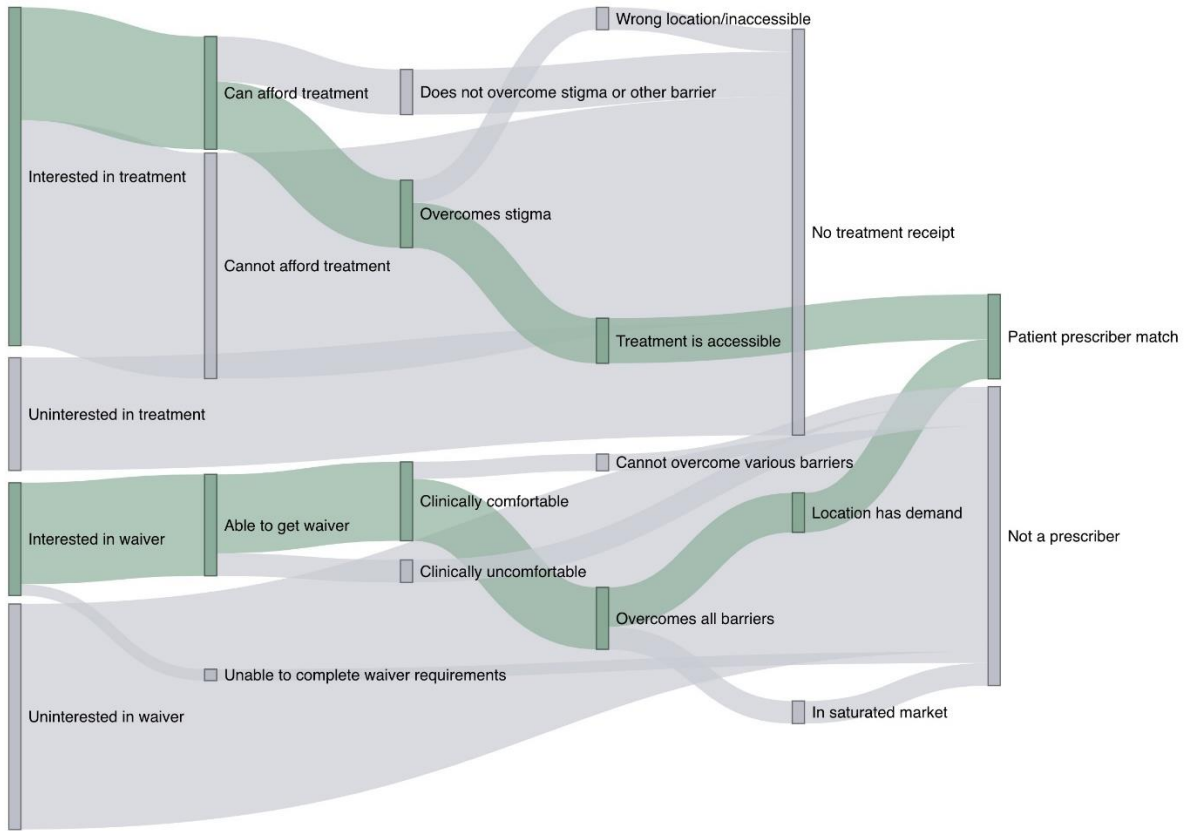


Figure S5. Treatment engagement as a dual continuum of care. Demand (treatment seekers) and supply (treatment providers) need to match in space and time for successful treatment entry, but both providers and seekers face numerous barriers along the way.

We assume that only people with use disorder will engage in MOUD treatment, as those without use disorder can simply voluntarily cease their drug use. Not all people with use disorder perceive a need for treatment or are interested in MOUDs. The hazard rate for people with use disorder making an effort to seek MOUD treatment is the *Tx seeking rate...* ($\rho_{T(\cdot)}$). Treatment-seeking can be thought of as attempting to inquire with a provider or program about receiving MOUD, regardless of whether MOUD is ultimately received.

Of those thus seeking treatment, some fraction will fail to receive it due to barriers such as affordability, acceptability, or stigma (*Tx seeking barrier loss fraction*, F^L). Estimates of this loss fraction are detailed in S3.c.ii.(2). The remainder are those who will engage in treatment as long as they have access to it (*Tx demand...*, $r_{UT(\cdot)}^*$):

$$r_{UT(\cdot)}^* = \rho_{T(\cdot)}(1 - F^L)U_{(\cdot)} \quad (2.13)$$

Treatment demand is then compared with treatment capacity to determine what fraction of demand can actually be met. We represent treatment capacity explicitly in the model, detailed below. If capacity is insufficient, that means some people will be unable to access treatment despite facing no other barriers to engagement:

$$r_{UT(\cdot)} = \text{MIN}(r_{UT(\cdot)}^*, K_{I(\cdot)}) \quad (2.14)$$

Where $K_{I(.)}$ is the *Tx intake capacity* at any given time, detailed below.

S2.d.ii.(2) Treatment capacity

Treatment capacity reflects the total number of patients nationwide who could be actively receiving a given treatment at any given time (*Tx capacity effective*, $K_{(.)}$). We calculate $K_{(.)}$ separately for each MOUD (subscripts B, M, V), but not each disorder type.

The maximum number of people who can be *in* treatment at a given time is distinct from the maximum number who can *enter* treatment, i.e., the maximum rate of treatment engagement (*Tx intake capacity*, $K_{I(.)}$); the latter depends on how much of existing capacity is already utilised, the rate of patients leaving treatment ($r_{TR(.)}$ and $r_{TU(.)}$), and the processing time required for someone seeking treatment to start receiving it (*Tx intake delay*, $\tau_{I(.)}$), which for simplicity we estimate at 1 month (0.083 years) for all MOUDs:

$$K_{I(.)} = \text{MAX} \left(0, \frac{K_{(.)} - \sum_u T_{u(.)} + \sum_u (r_{TRu(.)} + r_{TUu(.)})}{\tau_{I(.)}} \right), \quad u \in \{R, O, H\} \quad (2.15)$$

National-level data on treatment capacity are unfortunately and surprisingly very sparse (see S3.c.iii)). The limitations of data availability significantly constrain the level of detail with which we can represent treatment capacity, particularly for methadone and Vivitrol treatment. For these two MOUD types, we calculate effective treatment capacity $K_{(.)}$ as a fraction (*Tx effective capacity fraction*, $F_{(.)}^T$) of estimated nominal or theoretical treatment capacity ($K_{(.)}^*$):

$$K_{(.)} = K_{(.)}^* F_{(.)}^T, \quad (.) \in \{M, V\} \quad (2.16)$$

The effective capacity fraction captures a number of possible reasons why treatment capacity may not be fully utilised even in the face of demand, such as imperfect matching between demand and capacity due to geographic and temporal heterogeneity, or possibly treatment providers' and facilities' preferences for maintaining some capacity buffer.

We represent effective buprenorphine treatment capacity (K_B) in more detail, using data on the number of providers waived to prescribe buprenorphine (see S3.c.iii)). While the DATA 2000 buprenorphine waiver requirement and its different levels (38) create a certain theoretical maximum number of patients who could be receiving buprenorphine nationwide, in practice, providers face numerous other barriers to prescribing buprenorphine besides the waiver requirement, and rarely prescribe up to their full waived capacity (39, 40). We do not disaggregate these barriers, but they include factors like low reimbursement, lack of training, stigma, or lack of coordinating providers for, e.g., mental health services (41).

Empirical evidence indicates the average number of buprenorphine patients per provider increased initially, but has been decreasing for several years now (see S3.c.iii)). This pattern results from the combination of two trends. First, there is an overall trend of diminishing marginal returns to effective capacity from additional waived providers. This diminishing trend interacts with an early increase in patients per provider over the first several years after buprenorphine was approved for use in OUD treatment.

The overall diminishing trend likely arises for two main reasons. First, there is self-selection among providers in who gets waived first (42). Those providers who got waived early on (in the order of waiver receipt) were more likely to be those for whom addiction treatment was a major focus of their

practice, or those with many patients who showed a need for treatment, and therefore more likely to dedicate more time and effort to prescribing. Conversely, those waived later on are less likely to be focused on addiction treatment and less likely to make much time and effort available for buprenorphine prescribing. Second, there is some geographic mismatch between supply of waived providers and demand for buprenorphine treatment (43, 44), which tends to worsen with more waived providers. A growing fraction of later-waivered providers are in areas where capacity is plentiful and demand is already saturated, even while other locales still have unmet demand.

The early increases in patients or effective capacity per provider result primarily from exogenous changes in the waiver requirement itself. The DATA 2000 waiver allowed providers to increase their prescribing limits after a year (38, 45) and a 2005 amendment raised limits for group practice settings (46), allowing more capacity to come online gradually. In addition, there was likely some degree of learning as providers established and developed the practice-management infrastructure to handle this new treatment option, leading to gradual capacity growth.

To reflect the underlying trend of diminishing returns, we first calculate an indicated effective buprenorphine capacity K_B^* as the integral of an exponential decay function representing each additional provider's diminishing contribution to capacity (\hat{K}):

$$K_B^* = \int \hat{K} dB \quad (2.17)$$

$$\hat{K} = \hat{K}_0 e^{-\lambda_B B} \quad (2.18)$$

$$K_B^* = \frac{\hat{K}_0 e^{-\lambda_B B} + \hat{K}_0}{-\lambda_B} \quad (2.19)$$

Where B is the number of waived *Bup providers*, \hat{K}_0 is the initial or base effective capacity per provider (*Bup effective capacity per provider base*), and λ_B is a decay constant (*Bup effective capacity decay constant*) indicating the rate at which capacity added per additional provider diminishes. The effect of these parameters on the marginal effective capacity per new provider \hat{K} is shown in **Figure S6**.

Rather than represent in detail the policy changes and learning processes driving the early growth in patients per provider, we approximate these processes with a multiplier (F_{KB}) that adjusts from 0-100% of indicated effective capacity with an exponentially distributed delay:

$$K_B = K_B^* F_{KB} \quad (2.20)$$

$$F_{KB} = \int_{t \geq t_B} \frac{1}{\tau_{KB}} e^{-\frac{t-t_B}{\tau_{KB}}} dt = \begin{cases} 1 - e^{-\frac{t-t_B}{\tau_{KB}}}, & t \geq t_B \\ 0, & t < t_B \end{cases} \quad (2.21)$$

Where τ_{KB} is the *Bup effective capacity rampup time* and t_B is the *Bup rampup start year* in which capacity starts to come online.

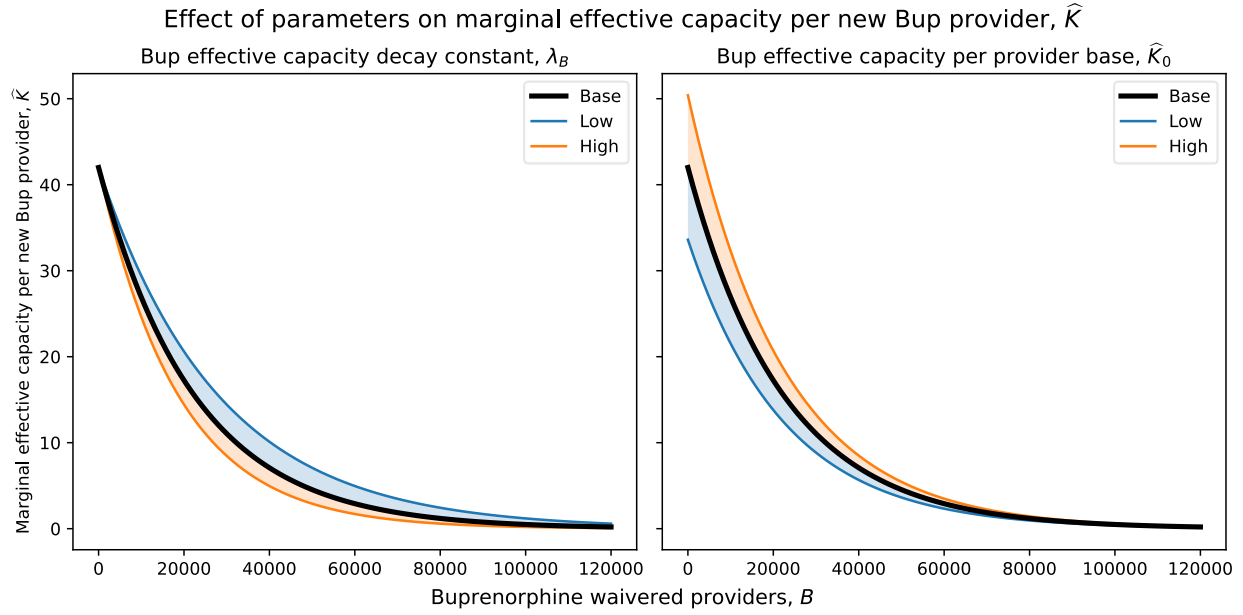


Figure S6. Functional relationship between buprenorphine-waivered providers and marginal effective capacity added per new waived provider. Panels show the effect of varying parameters λ_B and \hat{K}_0 by $\pm 20\%$. Note that while marginal capacity declines rapidly with additional providers, average capacity per provider never declines, as additional marginal providers only ever add capacity, never reduce it (see S3.c.iii).

S2.d.ii.(3) Treatment effects and outcomes

Patients in treatment will exit that state after a certain *Tx average duration* ($\tau_{T(.)}$) for each treatment type. Weighted averages for each MOUD were derived from an extensive review of literature; see S3.c.iv.(1).

Following treatment, patients exit to either a remission state ($r_{TR(.)}$), or back to use disorder ($r_{TU(.)}$). The proportion exiting to remission rather than back to use disorder (*Tx success fraction*, $p_{(.)}^R$) is itself a function of duration in treatment:

$$p_{(.)}^R = \frac{r_{TR(.)}}{r_{TR(.)} + r_{TU(.)}} = f(\tau_{T(.)}) \quad (2.22)$$

$$f(\tau_T) = \begin{cases} \left(\frac{\kappa_R^2}{1 + \kappa_R^2} \right) e^{\left(\frac{\lambda_R}{\kappa_R} \right) (\tau_T - m_R)} p^{RM}, & \tau_T \leq m_R \\ \left(1 - \left(\frac{1}{1 + \kappa_R^2} \right) e^{(-\lambda_R \kappa_R (\tau_T - m_R))} \right) p^{RM}, & \tau_T > m_R \end{cases} \quad (2.23)$$

The duration-success function, based on an asymmetric Laplace function, creates an asymmetric S-shaped curve (see **Figure S7**), whose shape and scale are based on a combination of expert judgment and existing studies (see S3.c.iv.(1)). The *Tx success fraction* function takes four parameters – the *inflection point* (m_R), asymmetry parameter *kappa* (κ_R), scale parameter *lambda* (λ_R), and the *max* possible success fraction (p^{RM}).

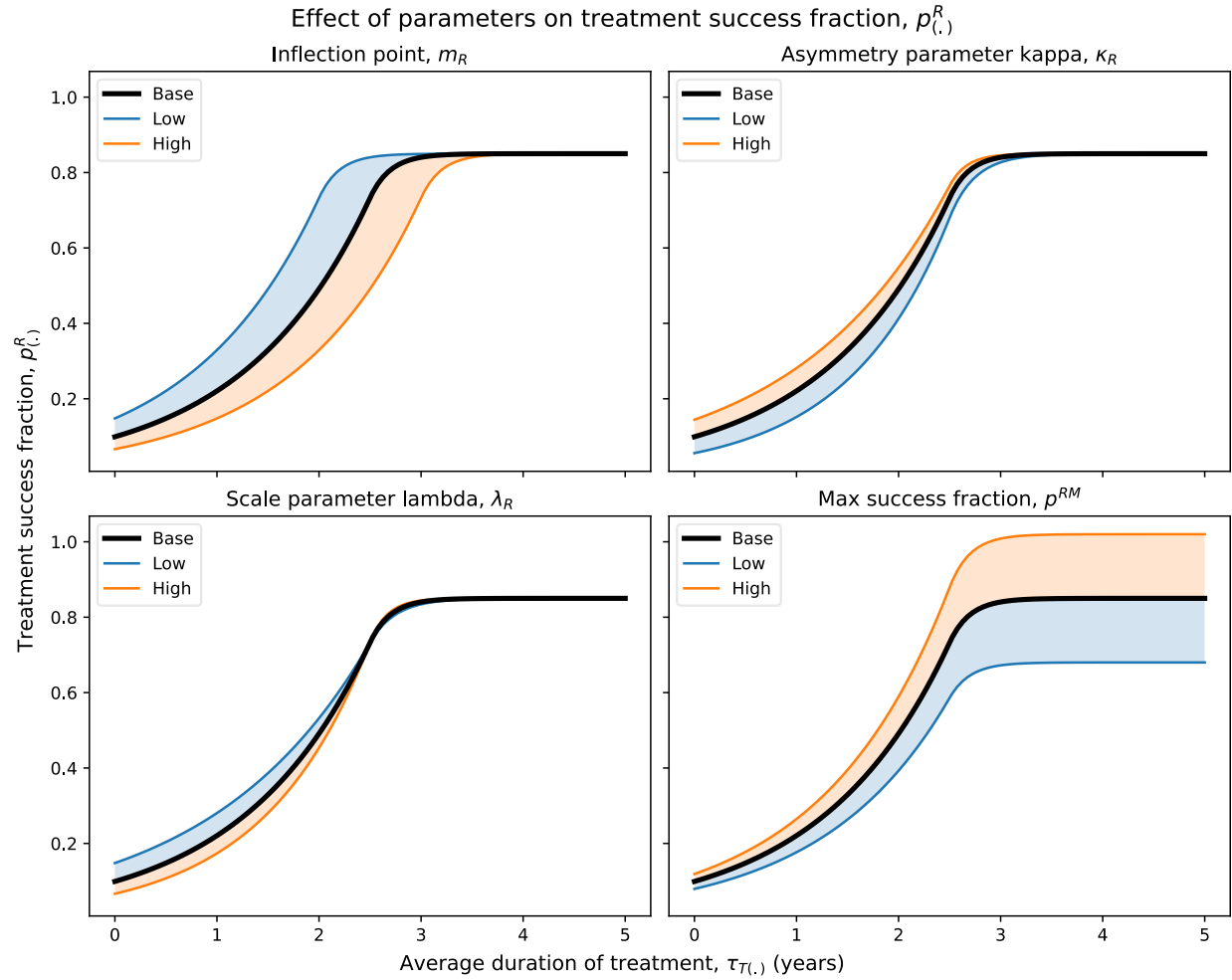


Figure S7. Functional relationship between duration of treatment and treatment success fraction. Panels show the effect of varying parameters m_R , κ_R , λ_R and p^{RM} by $\pm 20\%$.

Patients in the treatment stocks $T_{(.)}$ include a mix of people in one-year remission and people with ongoing use disorder. Transitions in and out of remission while in treatment are not uncommon, but to our knowledge there has been no attempt made to quantify these transition rates. For each treatment type, therefore, we specify a *remission fraction in Tx* ($F_{(.)}^R$), which is the average proportion of patients in that type of treatment whose use disorder is in remission. The fraction of treatment patients who are *not* in remission count towards the total number of people with use disorder, even though they are not in the $U_{(.)}$ stocks. For simplicity, we assume the remission fraction in each treatment type is equal to the current success fraction for that type:

$$F_{(.)}^R = p_{(.)}^R \quad (2.24)$$

Overdose and non-overdose death rates for treatment stocks ($\omega_{T(.)}$ and $n_{T(.)}$) are weighted by this fraction, with the portion of treatment patients in remission experiencing non-overdose deaths at the same rate as people in remission stocks ($R_{(.)}$) rather than use disorder stocks ($U_{(.)}$), and experiencing no overdose deaths. For those in treatment but not yet in remission, being in treatment nonetheless has beneficial effects on overdose and non-overdose death rates (*effect of MOUD Tx on OD death rate / non-OD death*

rate, $w_{(.)}^{TO} / w_{(.)}^{TN}$). Magnitudes of these effects are based on extant literature (see S3.c.iv.(2)). The net overdose death rate for a given stock of people in treatment is thus:

$$\omega_{T(.)} = \omega_{U(.)}(1 - F_{(.)}^R)w_{(.)}^{TO} \quad (2.25)$$

While the non-overdose death rate is:

$$n_{T(.)} = n_{U(.)}(1 - F_{(.)}^R)w_{(.)}^{TN} + n_{R(.)}F_{(.)}^R \quad (2.26)$$

Treatment also reduces opioid consumption for patients in treatment not yet in remission (*effect of MOUD Tx on Rx consumption*, $w_{(.)}^{Tq}$), thereby reducing their influence on demand for Rx opioids:

$$q_{DT(.)} = q_{DU(.)}(1 - F_{(.)}^R)w_{(.)}^{Tq} \quad (2.27)$$

S2.d.iii) Overdoses, naloxone, and synthetics

S2.d.iii.(1) Basic overdose death structure

The hazard rates of overdose and overdose death ($\omega_{(.)}$) differ based on drug use state. Overdose death data identify overdoses by the drug[s] involved (see S3.d.i)), but to keep the model estimation tractable, we instead allocate overdose deaths to user populations based on the populations' primary drug of use (Rx opioids vs. heroin), with further allocation of synthetic-opioid-involved deaths as detailed in S3.d.ii) below.

Not all overdoses result in death; sometimes death is averted through intervention, and sometimes an overdose is inherently less than lethal. For simplicity, we assume that the inherent lethality of an overdose and the probability that intervention occurs (or at least is attempted) are independent; many attempted interventions occur for overdoses that may not have resulted in death in the first place. In its basic form, therefore, we represent the overdose death rate ($\omega_{(.)}$) as:

$$\omega_{(.)} = \beta_{(.)}(1 - p_{S(.)})p_{D(.)} \quad (2.28)$$

The overall overdose rate ($\beta_{(.)}$) is multiplied by the complement of some base probability that overdoses are nonlethal, differentiating between Rx opioid and heroin overdoses (*base survival probability Rx OD / H OD*, p_{SR} / p_{SH}), and the probability that some lifesaving intervention does not successfully occur (*probability OD death not averted Rx / heroin*, p_{DR} / p_{DH}). We assume $p_{S(.)}$ is on average constant for a given substance, reflecting its inherent lethality given its usual modes of use; $p_{D(.)}$ is detailed further below.

Each use state (M, N, U_R, U_O, U_H) has its own base overdose rate parameter $\beta_{(.)}^*$, reflecting the combined effects of not only the substance involved and its usual modes of use, but also of frequency and patterns of use for that use state. For Rx OUD without heroin use (U_R), we also estimate a baseline (i.e., pre-illicitly manufactured fentanyl) synthetic-involved overdose rate (*overdose rate synth baseline*, β_R^S). We use the synthetic-involved overdose rate to help distinguish the effects of illicitly manufactured fentanyl from that of misused prescription fentanyl on overdose deaths, as detailed in S2.d.iii.(3) and S3.d.ii). Base overdose rates ($\beta_{(.)}^* / \beta_R^S$) and survival probabilities (p_{SR}^* / p_{SH}^*) are estimated model parameters.

S2.d.iii.(2) Intervention probability structure & naloxone probabilities

For a death to not be averted, none of the potential interventions that could prevent it can occur. An intervention can only occur if an overdose is first witnessed by someone who could intervene. For simplicity, we treat potential interventions as independent conditional on an overdose being witnessed, such that $p_{D(\cdot)}$ is the joint probability that none of them occur:

$$p_D = 1 - p_W p_I \quad (2.29)$$

$$p_I = 1 - \prod_j (1 - p_{Ij}) \quad (2.30)$$

Here p_W is the *probability OD witnessed* and p_{Ij} is the probability that intervention j successfully occurs, given that an overdose is witnessed. The value of p_W is derived from existing studies; see S3.d.v).

We represent two types of intervention, each with distinct probabilities of occurrence – bystander naloxone administration or calling emergency services. The *probability of calling emergency services* (p_{IE}) is a constant value estimated from literature (see S3.d.v)), which we assume results in a life-saving response by emergency medical services (EMS).

The probability of bystander naloxone administration (*probability Nx bystander...*, $p_{IB(\cdot)}$) depends on the amount of naloxone distributed. Specifically, we represent $p_{IB(\cdot)}$ using as a cumulative exponential distribution function of the density of naloxone kits distributed in the population:

$$p_{IB(\cdot)} = 1 - e^{-\lambda_N v(\cdot)} \quad (2.31)$$

Where λ_N is the *Nx kit distribution efficiency*, reflecting how effectively kits distributed end up in the times and places where they are needed, and $v(\cdot)$ is the number of *Nx kits per 100k population* for heroin or Rx users. λ_N is an estimated parameter; the effect of varying efficiency on $p_{IB(\cdot)}$ can be seen in **Figure S8**.

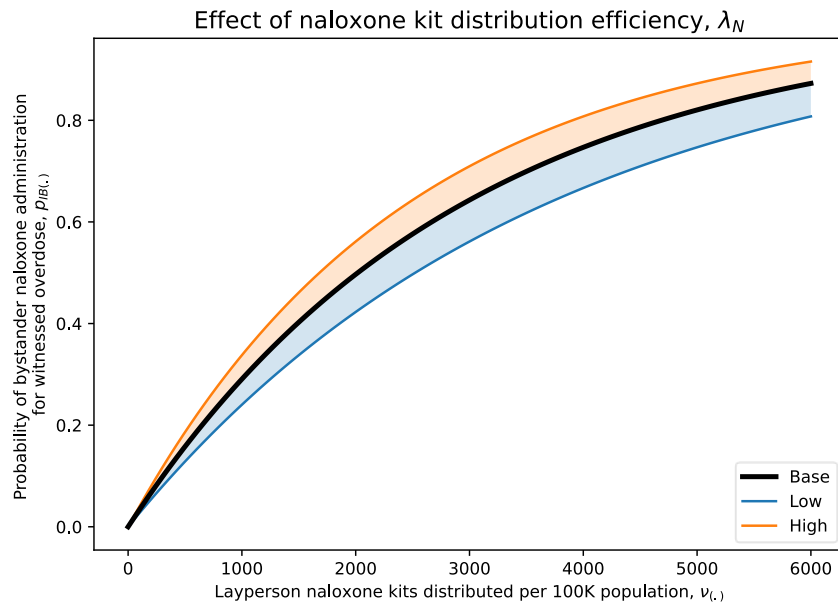


Figure S8. Functional relationship between naloxone kits distributed and probability of naloxone administration in the event of witnessed overdose, as dependent on naloxone kit distribution efficiency parameter λ_N .

Note that $v_{(.)}$ is normalised per 100,000 *total* people, rather than just people who use Rx opioids or heroin. The distinction between v_R and v_H depends on the channels by which naloxone is distributed and the populations such distribution focuses on (see S3.d.iv.(3)). The total amount of naloxone distributed is an exogenous time-series input, which is apportioned between Rx and heroin users based on an estimated parameter, the *fraction Nx kits to H users* (F^{NH}) (see S3.d.iv.(2)).

S2.d.iii.(3) Fentanyl effects on OD rates, survival rates, intervention probabilities

The prevalence of illicitly manufactured fentanyl in the illicit drug supply has increased rapidly since around 2013 (47, 48). Fentanyl is far more potent than other Rx opioids or even heroin (49), and has substantially affected overdose risks. We operationalise the effects of fentanyl on each stage of the overdose process based on its prevalence in the drug supply. Specifically, we drive the underlying growth in fentanyl prevalence with exogenous data (*fentanyl penetration curve*, ϕ ; see S3.d.iii)), representing the penetration of illicitly manufactured fentanyl in the heroin supply.

Note that while there is some evidence of fentanyl in counterfeit prescription opioids, especially on the west coast (23, 25, 27, 29), to our knowledge no quantitative data tracking counterfeit prevalence exists. We therefore cannot quantitatively account for illicit fentanyl in the Rx supply at this time, nor the effects of counterfeit prescription pills containing fentanyl on the street availability of Rx opioids.

Fentanyl penetration ϕ can be thought of as the average probability of exposure to fentanyl for users of heroin in any given instance of drug use. We can therefore approximate the average effects of fentanyl on overdose rates and survival probabilities as averages weighted by ϕ_H of their baseline heroin values and their corresponding values for fentanyl overdoses:

$$\beta_{(.)} = \beta_{(.)}^* (1 - \phi) + w_{\beta F} \beta_{(.)}^* \phi \quad (2.32)$$

$$p_{SH} = p_{SH}^* (1 - \phi) + p_{SF} \phi \quad (2.33)$$

Where $w_{\beta F}$ is the *fentanyl effect on OD rate H max*, i.e., how many times more likely overdose events are for fentanyl use relative to heroin use, and p_{SF} is the *base survival probability* of a fentanyl-involved overdose. These parameters, as well as the fentanyl penetration scaling factor ($s_{\phi H}$), are estimated in the main model calibration process.

Comparing overdose death rates against the counterfactual base death rates calculated using the base overdose rates and survival probabilities that exclude the effect of fentanyl allows us to attribute a certain portion of heroin-user deaths to the effects of illicit fentanyl (*overdose death rate synth...*, $\omega_{(.)F}$):

$$\omega_{(.)F} = \beta_{(.)} (1 - p_{SH}) p_{DHJ} - \beta_{(.)}^* (1 - p_{SH}^*) p_{DH} \quad (2.34)$$

We use this calculated death rate to estimate the contribution of illicit fentanyl penetration to overall overdose deaths.

S2.d.iii.(4) Nonfatal overdoses

We explicitly track nonfatal overdoses for each of the five main use states (M, N, U_R , U_O , U_H). The nonfatal overdose rate for each use state ($\eta_{(.)}$) is simply the difference between the overdose rate and overdose death rate for that state:

$$\eta_{(.)} = \beta_{(.)} - \omega_{(.)} \quad (2.35)$$

S3) Data Sources

S3.a) Main drug use states & transitions

Most data on drug use states and transitions in the model are drawn from the National Survey on Drug Use and Health (NSDUH). NSDUH allows us to distinguish individuals by the substances they have used in the past year (Rx opioids vs. heroin), as well as the degree of use associated with each substance (non-use vs. misuse / non-disordered use vs. use disorder). Both severity and substance of use are important distinctions on multiple dimensions – behaviourally, socially, clinically, and in terms of overdose and non-overdose risks. In particular, there is a clear psychological and social distinction between using prescription opioids and heroin (4, 5, 7, 50), as well as clear differences in both overdose risk and aggregate patterns of overdose mortality (51), which makes it crucial to distinguish between substances of use (but see also S2.b), footnote 1).

With two substances with three use states each, this creates a 3 x 3 matrix with 9 cells, of which 8 (excluding non-use of both) collectively map on to the 5 main drug use states in the model (M, N, U_R , U_O , U_H ; see **Figure S9** and **Table S8**), in combination with the fraction of people in treatment not in remission (see S2.d.ii.(3)). Broadly, we aggregated matrix cells based on what substance is associated with the highest severity of use disorder and/or risk of overdose.

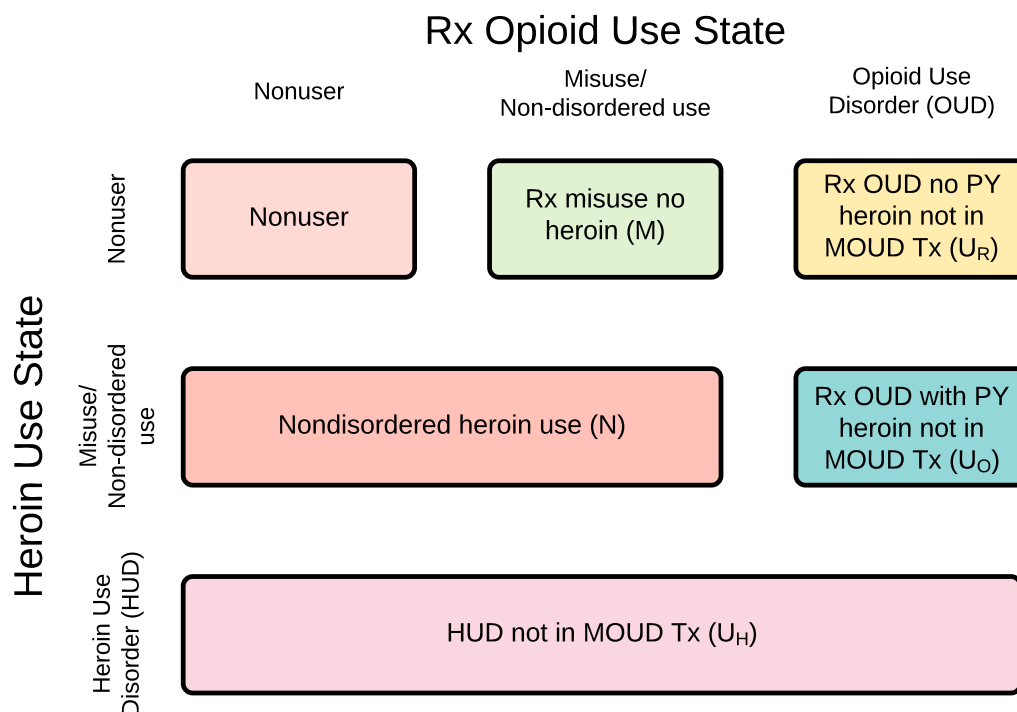


Figure S9. Prescription opioid / heroin use state matrix with corresponding NSDUH data variables or model states

Note that we have utilized preliminary NSDUH data for 2020 where available. Due to the COVID-19 pandemic, NSDUH sample sizes are somewhat smaller for 2020 than previous years; in addition, data from the NSDUH Restricted Data Analysis System are not yet available at time of writing. Despite these

limitations, we have incorporated these data to better reflect the most up-to-date trends and the effects of the COVID-19 pandemic on the opioid crisis (see also S6.d)).

S3.a.i) Prescription opioid misuse

We define Rx opioid misuse as including *any* use of someone else’s opioid prescription *or* use of Rx opioids solely “for the feeling [they] caused”. This definition matches the pre-2015 question wording in NSDUH (2).

Our definition of Rx opioid misuse excludes people sometimes called “medical misusers”, i.e., people who 1) used Rx opioids that were prescribed to them; 2) used them in ways other than as directed by their medical care providers; but 3) did so to treat pain (which is the intended therapeutic use of Rx opioids) and not for any other reason. This can include, for instance, using Rx opioids prescribed to oneself, in therapeutic doses, to treat pain, of the same kind for which they were originally prescribed, but without first consulting a medical professional regarding the repeat use.

S3.a.i.(1) Adjustments for NSDUH question change

From 2015 onward, NSDUH defines misuse more broadly, in a way that includes “medical misusers” or more specifically anyone who has used Rx opioids in *any* way not as directed by their medical care providers (2).

We account for this definitional change, which SAMHSA considers a trend break (2), using a fixed effect for the percentage increase in reporting of misuse due to the definitional change (*NSDUH misuse redefinition fixed effect, F^M*). This fixed effect parameter is estimated as part of the model calibration process (see S4)). From 2015 onward, we adjust the time series on Rx misuse and misuse initiation (see S3.a.v)) accordingly:

$$M_{adj}^y = \frac{M^y}{1 + F^M} \quad (3.1)$$

$$r_{MI_{adj}}^y = \frac{r_{MI}^y}{1 + F^M} \quad (3.2)$$

This adjustment reduces the number of people misusing after 2015 by approximately a third.

S3.a.ii) Prescription opioid use disorder

We define use disorder states according the DSM-5 criteria; however, NSDUH does not use the DSM-5 definition. Instead, we approximate the DSM-5 definition from NSDUH using the count of DSM-IV criteria for substance abuse or substance dependence that they meet. We ignore reported legal problems, which is no longer a DSM-5 criterion for disorder, and we are unable to include craving, which was added to the DSM-5 criteria but is not queried in NSDUH (see **Table S2**) (52). Note that our use of NSDUH’s DSM-IV criteria to approximate DSM-5 criteria differs from how NSDUH commonly reports ‘use disorder’ – NSDUH typically reports the union of DSM-IV substance abuse and substance dependence as ‘use disorder’, even though that more accurately reflects DSM-IV diagnoses rather than DSM-5 use disorder. This difference results in our calculated estimates for use disorder, particularly Rx OUD, being higher than what NSDUH reports as “use disorder.”

We separate people with Rx OUD who have not used heroin in the past year vs. those who have (U_R vs. U_O , see **Figure S9**). We adjust all NSDUH heroin-use estimates, including the count of people with Rx OUD with past-year heroin use (U_O^y), to account for systematic under-reporting (see S3.a.vii)).

The NSDUH counts of people with past-year use disorder also include that fraction of people in treatment states in the model ($T_{(i)}$) who are not yet in remission, who by definition have qualified for use disorder within the past year (see **Table S8**).

Table S2. DSM-5 criteria for substance use disorder, compared to DSM-IV substance abuse & substance dependence criteria

DSM-IV	Diagnostic criterion
*	Craving or a strong desire to use opioids
A	Recurrent opioid use resulting in failure to fulfill major role obligations at work, school, or home
A	Continued opioid use despite having persistent or recurring social or interpersonal problems caused or exacerbated by the effects of opioids
A	Recurrent opioid use in situations in which it is physically hazardous
D	Opioids are often taken in larger amounts or over a longer period of time than intended
D	There is a persistent desire or unsuccessful efforts to cut down or control opioid use
D	A great deal of time is spent in activities necessary to obtain the opioid, use the opioid, or recover from its effects
D	Important social, occupational or recreational activities are given up or reduced because of opioid use
D	Continued use despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by opioids.
D	*Tolerance, as defined by either of the following: (a) a need for markedly increased amounts of opioids to achieve intoxication or desired effect (b) markedly diminished effect with continued use of the same amount of an opioid
D	*Withdrawal, as manifested by either of the following: (a) the characteristic opioid withdrawal syndrome (b) the same (or a closely related) substance are taken to relieve or avoid withdrawal symptoms
DSM-IV column indicates DSM-IV diagnosis corresponding to DSM-5 criteria: A = substance abuse; 1 or more needed; 4 th criterion (legal problems) removed from DSM-5 D = substance dependence; 3 or more needed in 12-month period * Craving is a DSM-5 criterion not included in DSM-IV diagnoses and not queried in NSDUH Source: (52)	

S3.a.iii) Non-disordered heroin use

Approximately one-quarter of the people who report heroin use in the past year in NSDUH do not meet use disorder criteria for their heroin use. Anyone who reports past-year heroin use in NSDUH who does not qualify for HUD is counted either as having non-disordered heroin use (N^y) if they do not qualify for Rx OUD, or Rx OUD with past-year heroin use (U_O^y) if they do (see **Figure S1**). We adjust all NSDUH heroin-use estimates, including the count of people with NDHU (N^y), to account for systematic under-reporting (see S3.a.vii)).

S3.a.iv) Heroin use disorder

As with Rx OUD, we approximate the DSM-5 use disorder definition using NSDUH criteria (see above). Note that substance use disorder associated with Rx opioid use vs. heroin use are sometimes both

collectively referred to as ‘opioid use disorder’, since heroin is an opioid substance. However, NSDUH queries each UD criterion for each substance separately, allowing us to identify whether UD is associated with use of Rx opioids, heroin, or both. For clarity, we use ‘Rx OUD’ to refer to UD associated with use of Rx opioids, and ‘HUD’ to refer to UD associated with use of heroin or both.

We adjust all NSDUH heroin-use estimates, including the count of people with HUD (U_H^y), to account for systematic under-reporting (see S3.a.vii)).

The NSDUH counts of people with past-year use disorder also include that fraction of people in treatment states in the model ($T_{(t)}$) who are not yet in remission, who by definition have qualified for use disorder within the past year (see **Table S8**).

S3.a.iv.(1) Caveat regarding HUD data

NSDUH’s 2018 data on HUD prevalence and heroin use initiation show a downward trend from previous years, and 2019 data continue this downward trend, showing a sharp decline. The drop is large and rapid enough that several subject-matter experts expressed concern about the accuracy of the data. Changes in overdose mortality and MOUD treatment engagement are insufficient to explain the drop, but without specific data on remission (and relapse), we cannot conclusively demonstrate the physical impossibility or inconsistency of the reported numbers.

In consultation with our subject-matter experts, we have considered several plausible explanations for the decline – increased under-reporting due to growing fear or stigma, possibly associated with fentanyl; increasing self-identification as a fentanyl rather than heroin user (e.g., in regions where fentanyl has almost completely displaced heroin); and decreasing relapse due to ‘older’ cohorts of former heroin users attaining an increasingly durable state of remission. We found no evidence for the first two of these explanations, and our subject-matter expert team considered them less likely than there being issues with the NSDUH data.

Increasing durable or sustained remission was the only other explanation supported by our subject-matter experts as well as existing literature. We modified the model’s remission structure (see S2.b) and S3.c.v)) to more accurately reflect this effect, which improved model performance but was insufficient to produce the observed decline.

We have queried SAMHSA directly about the 2019 HUD and heroin initiation data, and they report that the 2019 data are in no way anomalous as no methodological changes occurred that may account for the difference.

With no further explanation or justifiable alternative, we have estimated the model on the assumption that the NSDUH 2019 HUD and heroin initiation data are no less accurate than in other years. This has several implications for the model’s estimates, behaviour and projections.

Most importantly, the rapid drop in initiation indicates the risk response feedback (see S2.c.ii)) is very strong, exerting a dominant effect on the system in the last few years as the sharp rise in overdose mortality due to illicit synthetics deters new heroin initiates. Similarly, the fall in HUD prevalence indicates relatively high rates of remission vs. relapse, absent new initiations or use disorder development. With sustained high overdose mortality, this strong behavioural response results in a substantial projected decline in opioid use and mortality over the next decade (see S5.c)).

If the 2019 heroin use data turn out to be, for whatever reason, a substantial under-estimate, then our model will have over-estimated the strength of this risk response feedback, as well as rates of HUD remission vs. relapse. A weaker risk response and lower remission / higher relapse rates will result in persistently higher levels of opioid use and overdose mortality than we are currently projecting, with a much slower decline.

S3.a.v) Initiating prescription opioid misuse

We derive data on annual initiation rates of Rx opioid use from NSDUH's Restricted Data Analysis System (RDAS), which allows identification of past-year initiates. The data do not directly distinguish between initiation of use with vs. without a prescription (r_{MI} vs. r_{MD}). To make that distinction, we use the fraction of past-year initiates who report that the source of Rx opioids for their *most recent* instance of misuse was one or more of their own prescriptions (vs. other sources), averaged over time, as a proxy for the fraction initiating misuse from a prescription. Due to the trend break in misuse reporting in 2015, we use separate fractions before 2015 and from 2015 onward.

S3.a.vi) Initiating heroin use

We derive data on annual initiation rates of heroin use from NSDUH's Restricted Data Analysis System (RDAS) as well, using it to identify whether individuals are initiating heroin with no past-year Rx opioid use (r_{ND}), with past-year Rx opioid misuse (r_{MN}), or with past-year Rx OUD (r_{UO}). These data were then adjusted to address under-reporting, as outlined below.

S3.a.vii) Heroin use adjustments

NSDUH estimates of the number of people who use heroin are notoriously low (53–58). This under-reporting is due in part to exclusion of incarcerated populations where heroin use is disproportionately common, and in part to the strong stigma associated with heroin use. To correct for this under-estimation, we adjust all NSDUH data on prevalence and initiation of heroin use (N^y , U_O^y , U_H^y , r_{ND}^y , r_{MN}^y , r_{UO}^y) as follows.

No adjustment to empirical data should ever be undertaken lightly. We make this change noting that 1) the systematic problems with the data are well-known, and 2) the alternative of not adjusting the data would be worse, forcing skewed estimates of various parameters and creating errors that would propagate throughout the model (due to its enforced internal consistency and conservation of matter). Our adjustments are based on extensive literature review as well as discussions with subject matter experts.

We base the adjustment on estimates of chronic heroin users (CHU) from the RAND Corporation report “What America's Users Spend on Illegal Drugs, 2006-2016” (56, 59). The report estimates number of CHUs for 2006-2016. We compare the RAND CHU population against the total NSDUH reported population of heroin users year by year ($N_t^y + U_{Ot}^y + U_{Ht}^y$), yielding an average ratio of 3.11. We then multiply each NSDUH heroin use population and initiation flow by this ratio. Note that the actual ratios of RAND to corresponding NSDUH estimates in the data decline over time, but for simplicity, we use a single average figure for each population group, meaning the temporal trends in the data are driven by NSDUH. This may result in some underestimation of heroin users in our adjusted data in earlier years, and some overestimation in later years.

S3.a.viii) Total population

As great a problem as the opioid crisis may be, the total U.S. population is orders of magnitude larger. We therefore represent total population – or more accurately, the NSDUH survey population of non-institutionalised individuals aged 12 and older – as exogenous.

Total population figures for 1999-2020 are taken directly from NSDUH data. Projected total population for 2020-2032 is calculated using the ratio of the 2020 NSDUH population to 2020 population projected by the U.S. Census. This ratio (approx. 83%) is applied to U.S. Census projected population for 2020-2032 to yield a projection of the NSDUH population:

$$\text{Projected NSDUH pop.} = \frac{2019 \text{ NSDUH pop.}}{2019 \text{ US Census pop.}} \times \text{Projected US Census pop.} \quad (3.3)$$

S3.b) Opioid prescribing, supply, and price data

S3.b.i) Prescription opioid supply

We draw most data regarding prescribing patterns, as detailed in S2.d.i.(1), from several proprietary IQVIA datasets – the National Sales Perspective®(NSP), National Prescription Audit® (NPA), and Total Patient Tracker® (TPT). NPA and TPT are national-level projected services designed to estimate the total number of prescriptions dispensed and unique (non-duplicated) patients receiving prescriptions within a specified timeframe respectively, across all drugs and therapeutic classes. NPA captures prescriptions dispensed in the outpatient setting at U.S. retail and mail-order pharmacies, as well as pharmacies that dispense to long-term care facilities, while TPT projects patient counts based on prescriptions dispensed from U.S. retail pharmacies. TPT uses prescription activity as part of their projections and integrates information from pharmacies to eliminate duplicate patients. As of 2019, IQVIA data captures 92% of all dispensed prescriptions in the U.S. from a sample of about 49,900 retail pharmacies, for a total of over 3.5 billion transactions annually. The prescription coverage and sample size have varied across the time period in our analyses. NPA and TPT are projected to the known universe of retail pharmacies.

NSP estimates the volume of prescription drug products moving from distributors and manufacturers into various retail and non-retail outlets, in terms of sales dollars and product quantities. Retail outlets include various pharmacy settings, including mail-order; non-retail outlets include clinics, hospitals, long-term care facilities, and other such settings. NSP captures 86% of sales in the retail channel and 97% of the sales in the non-retail channel, or about 90% of the U.S. pharmaceutical market in total. It includes sales from 388 indirect suppliers and direct sales reported from around 100 manufacturers, totaling about 1.5 billion transactions per year.

S3.b.i.(1) Total prescriptions and MMEs

For the total number of opioid analgesic prescriptions dispensed annually (*total prescription opioid Rx*), we use IQVIA NPA® data on the total number of prescriptions for all opioid analgesic products dispensed from U.S. outpatient pharmacies (retail and mail-order).

We calculate the total annual MMEs (*total Rx MME prescribed*) by multiplying the opioid units (e.g., tablets, patches, liquid volume) reported in IQVIA NPA by the milligram strength per unit and by the appropriate MME conversion factors as shown in **Table S3**.

Table S3. MME conversion factors for different opioid substances

Opioid	Dosage Formulation	MME conversion factor	Opioid	Dosage Formulation	MME conversion factor
Benzhydrocodone	Oral solid/liquid	1.22	Meperidine	Injectable	0.3
Buprenorphine	Injectable	100	Meperidine	Oral solid/liquid	0.1
Buprenorphine	Oral solid/liquid	0.03	Methadone	Injectable	4.29
Buprenorphine	Patch	12.6	Methadone	Oral solid/liquid	3
Butorphanol	Injectable	15	Morphine	Injectable	3
Butorphanol	Nasal spray	7	Morphine	Oral solid/liquid	1
Codeine	Oral solid/liquid	0.15	Morphine	Rectal	1
Dihydrocodeine	Oral solid/liquid	0.25	Nalbuphine	Injectable	3
Fentanyl	Injectable	150	Opium	Oral solid/liquid	1
Fentanyl	Oral solid/liquid	0.13	Oxycodone	Oral solid/liquid	1.5
Fentanyl	Patch	7.2	Oxycodone (Xtampza)	Oral solid/liquid	1.67
Fentanyl (Lazanda)	Nasal spray	1.28	Oxymorphone	Injectable	30
Fentanyl (Subsys)	Nasal spray	0.18	Oxymorphone	Oral solid/liquid	3
Hydrocodone	Oral solid/liquid	1	Pentazocine	Injectable	1
Hydromorphone	Injectable	20	Pentazocine	Oral solid/liquid	0.37
Hydromorphone	Oral solid/liquid	4	Tapentadol	Oral solid/liquid	0.4
Hydromorphone	Rectal	4	Tramadol	Oral solid/liquid	0.1
Levorphanol	Oral solid/liquid	11			
Sources: (107–110)					

Note that neither total prescriptions nor total MMEs is used directly in the model; instead they are combined to yield the *avg MME per opioid Rx*, which together with the *average days per prescription* (m_D , see S2.d.i.(1)) is used to calculate the *average MME per day* (m_M). They are also used to derive a number of other prescribing-related time series as explained below.

There are recognized limitations with using MME as a standardising conversion factor for quantifying the potency of opioids across different opioid moieties. MME conversion cannot fully account for pharmacologic variability due to moiety, route of administration, and patient characteristics; furthermore, conversion factors are inconsistently applied and interpreted in clinical practice and in research (60, 61). These limitations likely mean that some moieties are under- or over-represented in total calculations of the Rx supply.

However, in no place does SOURCE directly make use of calculated MMEs or MME per prescription and daily dosage values. Instead, these values are either normalised to their initial values (see S2.c)) or used to calculate fractions (see S3.b.i.(4)), and it is the temporal trend in these normalised or fractional values that drives changes in the model. As such, barring extreme shifts in the composition of overall opioid prescribing by opioid moiety, the exact MME conversion factors used in calculating total MMEs makes little difference to the overall temporal trends in prescribing and hence to the behaviour of the rest of the model. The model also does not rely on MME to estimate overdose or other hazard rates.

S3.b.i.(2) Total patients receiving prescriptions

For the total number of *patients receiving opioid prescription annual* each year (m_P) we use IQVIA TPT® data on the total unique patients receiving opioid analgesic prescriptions within each year from 2002-2020. Prior to this period, we calculate the value for m_P using simple linear extrapolation of the TPT® annual patient totals.

S3.b.i.(3) Prescriptions per person and average duration

For the average number of *prescriptions per person* (m_N) over time, we use data from IQVIA TPT® and NPA®, described above. The TPT® data provide estimates of total unique patients receiving opioid analgesic prescriptions within each year from 2002-2020, while the NPA® data provide estimates of total opioid analgesic prescriptions dispensed each year from 1999-2020 (see above). We calculate *prescriptions per person* (m_N) over time using the ratio of total prescriptions in NPA® to total patients in TPT® each year.

To calculate the *average days per prescription* (m_D), we first estimate the *average duration of medical opioid use* (τ_M) by comparing IQVIA TPT® data on unique patients receiving opioid analgesic prescriptions at different levels of temporal resolution. Specifically, we use the ratio of unique patients each month (*patients with current month opioid Rx*) reported in TPT® to unique patients each year as a proxy for duration of medical use. On average, at any given time, the fraction of all patients receiving prescriptions within the year whose prescriptions are currently active will be equivalent to the fraction of the year for which, on average, a patient's prescriptions are active. This allows the monthly to annual unique patients ratio to serve as a proxy for duration of active (or more accurately, current-month) medical use.

We then compare the *prescriptions per person* (m_N) with this estimate of *average duration of medical opioid use* (τ_M , converted to days) to yield the *average days per prescription* (m_D).

S3.b.i.(4) ADF fraction of prescribed supply

We calculate a time series for the *ADF fraction of prescribed Rx opioids* (F^{AR}) (see S2.d.i.(2)) using the same IQVIA NPA data and MME conversion factors used to calculate total annual MMEs above (**Table S3**). We use a list of all FDA-approved ADF opioids currently marketed in the United States (see **Table S4**) to identify the total annual MMEs prescribed for ADF products and divide that by *total Rx MME prescribed* to arrive at the ADF fraction (of MMEs prescribed) for each year.

Table S4. FDA-approved abuse-deterrent formulation opioids currently marketed in the U.S.*

FDA-approved ADF opioids					
Product name	Active ingredient	Year approved	Product name	Active ingredient	Year approved
Arymo™ ER	Morphine	2017	OxyContin®	Oxycodone	2010
Embeda®	Morphine	2009	RoxyBond™	Oxycodone	2017
Hysingla® ER	Hydrocodone	2014	Xtampza® ER	Oxycodone	2016
MorphaBond ER™	Morphine	2015			
* Three other abuse-deterrent formulations are approved but not marketed in the U.S., and are therefore excluded from this analysis: Targiniq™ ER, Troxyca® ER, and Vantrela™ ER					

S3.b.i.(5) OxyContin withdrawal street supply shock

We include a single historical Rx street supply shock (see S2.d.i.(2)), representing the August 2010 withdrawal of non-ADF OxyContin. To estimate the magnitude of this shock, in terms of the proportion of the street supply impacted, we used StreetRx, a crowdsourced database of street prices paid for illicit substances. StreetRx reports include information on substance, quantity, and price. We set the magnitude of the shock at 0.45 (where 100% of street supply = 1), equal to the fraction of total MMEs reported in StreetRx for 2010 consisting specifically of OxyContin (excluding other oxycodone). OxyContin MMEs are calculated based on conversion factors in **Table S3**.

S3.b.ii) Prescription opioid demand

To calculate the *Rx demand for misuse* (q_o), we use the number of people in each opioid use state multiplied by the per-person demand for opioid use for that use state, expressed in MMEs per year. We calculate per-person demand based on NSDUH data on average number of days of use per year, rounded to the nearest 10 days, reported over the 2010-2018 period: 50 days for M, 110 days for U_R , 180 days for U_O , 100 days for N, and 120 days for U_H . (The latter two categories are modified by the average fraction over 2010-2018 of people in those states who also use Rx opioids.) Note that these reported days of use are likely underestimates, particularly for the U_R and U_O groups. We multiply the days of use by assumed MME per day values of 40 MME/day for non-disordered groups (M and N), and 100 MME/day for use disorder (U_R , U_O , U_H). We believe these are conservative estimates, and actual use quantities are likely higher.

S3.b.iii) Heroin price

We calculate a normalised index of heroin price using data from two sources – U.S. wholesale prices for heroin from the UN Office on Drugs and Crime (62), and heroin retail prices from the DEA System to Retrieve Information from Drug Evidence (STRIDE), as used in (63). These two sources cover different years (2007-2018 vs. 2002-2011, 2013 & 2015 respectively). To combine them, we first normalised each price series to its 2007 value. We combined the two 2007-normalised indices, taking the mean in years where both were available and using whichever was available otherwise. Finally, we re-normalised the combined index to 1999, the year of model initialisation.

S3.c) Treatment & remission

SOURCE explicitly represents use of the three FDA-approved MOUDs – buprenorphine, methadone, and Vivitrol. Other forms of treatment (e.g., psychosocial, mutual aid group, etc.) are not explicitly represented; their effects are incorporated into non-MOUD remission pathways ($r_{UR(i)}$). Note that in contrast to drug use states, which represent *past-year* use, the treatment states in the model represent current, ongoing treatment receipt.

S3.c.i) Treatment receipt

We represent buprenorphine treatment receipt using IQVIA Total Patient Tracker® (TPT) data (see also S3.b.i)), which reports estimated total unique patients receiving buprenorphine within each month. This total includes only people receiving buprenorphine products designated for use as opioid antagonists (i.e., as MOUD) and not for pain. We use this current-month patient total as a proxy for the total patients receiving buprenorphine at any point in time (T_B^y).

Methadone maintenance treatment (MMT) receipt (T_M^y) is estimated using N-SSATS point-in-time counts as of March 31 for each year from 1999-2020, with data interpolated for missing years.

For Vivitrol receipt, we use IQVIA National Sales Perspective®, which reports annual injections of Vivitrol. Because Vivitrol can also be used for alcohol use disorder (AUD), and IQVIA does not report the indication for use, we subtracted the average number of injections from 2006-2010 (prior to Vivitrol’s approval for OUD treatment) from subsequent years, to arrive at estimates for Vivitrol injections for OUD. These estimates were then divided by 12, as injections are usually given monthly, to arrive at point-in-time counts for patients receiving Vivitrol (T_V^y).

S3.c.ii) Treatment-seeking and barriers

S3.c.ii.(1) Treatment seeking rates

We estimate a single base treatment-seeking rate in the model, which is the total treatment-seeking rate across MOUD types for people with Rx OUD without heroin use ($\rho_{TR} = \rho_{TRB} + \rho_{TRM} + \rho_{TRV}$), as part of the model calibration process. This base rate provides an anchor for all other treatment-seeking rates in the model.

In the absence of more detailed data to distinguish the states, we assume people with Rx OUD with heroin use seek treatment at the same rate as those without heroin use ($\rho_{TO(.)} = \rho_{TR(.)}$). Most literature on treatment does not distinguish between these two groups; indeed, most literature on treatment focuses on people with HUD rather than Rx OUD.

We express total HUD treatment seeking rate (ρ_{TH}) as a multiple of the base rate:

$$\rho_{TH} = m_{TRH}\rho_{TR} \quad (3.4)$$

Where m_{TRH} is the *Tx seeking rate HUD relative to Rx OUD no H*. We set $m_{TRH} = 4.84$, based on 2020 NSDUH data on the fraction of people with HUD reporting receipt of MOUD treatment compared to the equivalent fraction for people with Rx OUD (64).⁴

Treatment-seeking rates for each MOUD type are expressed as fractional multipliers of the total base rate, differing for Rx OUD vs. HUD:

$$\rho_{Tij} = m_{Tij}\rho_{Ti}, \quad i \in \{R, H\}, j \in \{B, M, V\} \quad (3.5)$$

The values of each of these fractions are based on expert estimates on relative patient preferences for each treatment type (see **Table S5**). For instance, while buprenorphine treatment is generally the most popular (65, 66), people with HUD are much more likely to seek MMT than people with Rx OUD.

Table S5. Rates of treatment-seeking by use disorder and MOUD type, relative to total Tx-seeking rate for OUD (r_{UTR})

	Total MOUD Tx-seeking rate	Buprenorphine	Methadone	Vivitrol
Opioid use disorder	0	0.625	0.05625	0.31875
Heroin use disorder	4.84	2.6637	1.743	0.4359

⁴ 2019 is the first year that NSDUH queries MOUD receipt specifically, as opposed to general treatment receipt. The calculated ratio of ~5 nonetheless accords with with consistent data from both NSDUH and TEDS indicating people with HUD seek or receive treatment at far higher rates than people with Rx OUD.

Note that we do not account for people switching between medications within a given treatment episode, though it is possible for someone who receives one MOUD during one treatment episode to subsequently receive a different one later. We also assume patients seek a specific MOUD during a given treatment-seeking attempt, as the three available medications are generally viewed as quite different.

S3.c.ii.(2) Barriers to treatment receipt

We calculate the *Tx seeking barrier loss fraction* (F^L), i.e., the fraction of people seeking treatment who fail due to barriers such as affordability, acceptability, or stigma, based on data from NSDUH. Specifically, for people who make an effort to get treatment but do not receive it, NSDUH offers 15 potential reasons for non-receipt. We divide these reasons into three categories (see **Table S6**) – 1) affordability, e.g., lack of health insurance or insurance that doesn’t cover treatment; 2) accessibility, e.g., lack of transportation to get to a treatment provider or providers not having space available for new patients; and 3) stigma and other non-affordability issues, e.g., fear of potential negative opinions or belief that treatment will not help.

Since the model explicitly represents treatment capacity constraints, which captures the loss of potential treatment patients due to accessibility reasons, we do not include those people who report non-receipt exclusively for accessibility reasons in F^L . Instead, we include only those who report at least one of affordability and stigma or other non-affordability issues as reasons for treatment non-receipt:

$$F^L = \frac{\text{Treatment seekers not receiving due to affordability or stigma barriers}}{\text{Total treatment seekers, receiving or not}} \quad (3.6)$$

Table S6. Barriers to treatment engagement queried in NSDUH (with variable codes) and corresponding overarching categories used in model

Barrier to treatment engagement	Variable code	Category
Need treatment but no health coverage or cannot pay	NDTRNNOCOV	Affordability
Need treatment but insurance doesn’t cover substance use treatment	NDTRNNOTPY	
Need treatment but transportation posed a difficulty	NDTRNTSPHR	Accessibility
Need treatment but the type desired is not available	NDTRNWANTD	
Need treatment but the treatment centers had no open spaces	NDTRNPFULL	
Need treatment but don’t know where to get it	NDTRNDKWHR	
Need treatment but afraid neighbors would have a negative opinion	NDTRNNDRNG	Stigma & other
Need treatment but afraid job will have a negative opinion	NDTRNJOBNG	
Need treatment but afraid others would find out	NDTRNFNDOU	
Need treatment but not ready to stop using	NDTRNNSTOP	
Don’t think you need treatment	NDTRNNONED	
Need treatment but think you can handle the problem without it	NDTRNHANDL	
Need treatment but don’t think that it will help	NDTRNNOHLP	
Need treatment but don’t have time	NDTRNNTIME	
Some other reason	NDTRNMIMPT	

We capture the effect of accessibility barriers through the treatment capacity constraint. The *Tx demand fulfilment ratio* reflects how much of treatment demand, after accounting for non-accessibility barriers, can be met given the available capacity:

$$DFR = \frac{r_{UT(.)}}{r_{UT(.)}^*} = \frac{\text{MIN}(r_{UT(.)}^*, K_{I(.)})}{r_{UT(.)}^*} \quad (3.7)$$

We calculate a prior (see S4.a.ii)) for this value for buprenorphine in 2018 ($\frac{r_{UTB}}{r_{UTB}^*}$) based on a recent audit study (67), which tracked treatment-seeking attempts and the success rate at obtaining an appointment for buprenorphine treatment. Specifically, we use the number of appointments offered as a proxy for r_{UTB} , and the sum of appointments offered and attempts failed due to access or capacity barriers as a proxy for r_{UTB}^* , yielding a calculated demand fulfilment ratio of 58.7% in 2018 (see S4.b)).

S3.c.iii) Treatment capacity

We calculate total buprenorphine-waivered providers (B), used to calculate effective buprenorphine capacity (K_B) as described in S2.d.ii.(2), using multiple literature sources (see S3.e)). These studies have reported the estimated number of buprenorphine (i.e., DATA 2000) waived providers each year since 2003, when buprenorphine was first approved for OUD treatment.

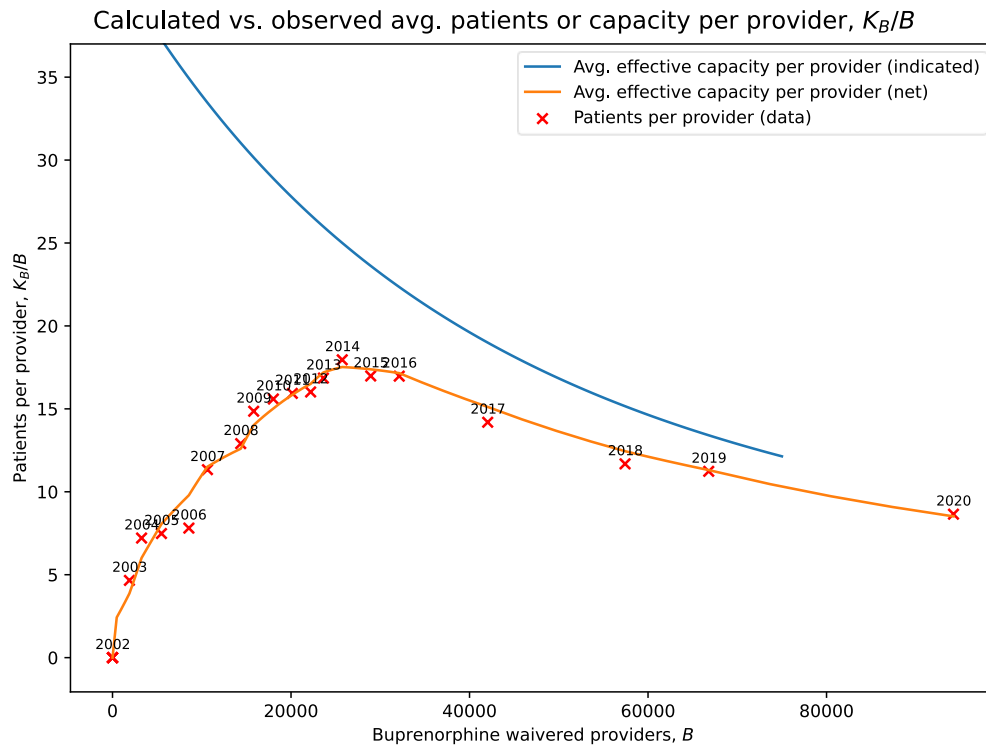


Figure S10. Observed patients per provider against number of waived providers in historical data, compared with indicated and net average effective capacity per provider in model. The indicated average effective capacity (blue) reflects the diminishing marginal capacity per waived provider, while the net average effective capacity (orange) also incorporates the effect of the capacity rampup delay.

Comparing the number of patients receiving buprenorphine against the number of waived providers shows that since around 2011, the average number of patients per waived providers has been dropping as more providers get waived (**Figure S10**). Combined with evidence that capacity continues to be a binding constraint on receipt of buprenorphine treatment (see above), we infer from this that the marginal contribution of each new waived provider to effective capacity is diminishing, as explained in

S2.d.ii.(2). We also use the patients-per-provider data to estimate the *Bup effective capacity rampup time* (τ_{KB}), which affects the fraction of indicated effective capacity per waived provider that is actually online. We estimate the *Bup effective capacity decay constant* (λ_B) at $4.45\text{E-}5$, base capacity per provider (\hat{K}_0) at 42, and τ_{KB} at 12 years, to match the empirically observed pattern (**Figure S10**; see also **Figure S6**). We define the *Bup rampup start year* at which point capacity starts to adjust toward indicated capacity as $t_B = 2000.0$ (note that this is actually prior to when buprenorphine was actually approved and the first providers waived to approximate the fact that initial waived providers had nonzero capacity at the time they were first waived and reflected in the data).

To our knowledge, there are no time series data available for methadone (MMT) and Vivitrol capacity (including from national treatment surveys such as N-SSATS). As such, we estimate theoretical capacity ($K_{(.)}^*$) based on the number of patients receiving each of these types of treatment (see S3.c.i)) divided by the capacity utilization percentages ($F_{(.)}^{TU}$) reported for each in N-SSATS:

$$K_{(.)}^* = \frac{T_M^y}{F_{(.)}^{TU}}, \quad (.) \in \{M, V\} \quad (3.8)$$

Absent additional data, we assume the effective capacity fraction is equal to the capacity utilization percentage ($F_{(.)}^T = F_{(.)}^{TU} = 0.866$ for methadone and 0.88 for Vivitrol). Note that this creates a circularity – effective treatment capacity ($K_{(.)}$) will be *exactly* equal to the number of patients receiving treatment $T_{(.)}^y$, resulting in an artificially perfect fit between simulated patient numbers and data as long as capacity is the binding constraint on treatment receipt in the model. This circularity can only be resolved with additional data on treatment capacity.

S3.c.iv) Treatment duration, outcomes, and effects

S3.c.iv.(1) Average treatment duration and outcomes

Average durations for buprenorphine, methadone, and Vivitrol treatment ($\tau_{T(.)}$) are calculated from mean or median reported durations of treatment, weighted by sample size, in multiple studies over two decades (see S3.e)), at 0.61 , 1 , and 0.22 years respectively. Where studies reported only medians but not means, we approximated the mean using the approach from (68) based on either the interquartile range or minimum and maximum.

While a positive relationship between duration of retention in treatment and ‘successful’ treatment outcomes (i.e., sustained remission either in or after leaving treatment) is well-established (69), we identified only one study that actually reports 1) what fraction of treatment patients leave treatment ‘successfully’ vs. return to use disorder, 2) at various durations of treatment, and 3) how long on average each subgroup of patients remains in treatment (70).⁵

As such, we also drew on expert estimates to quantify the relationship between duration and outcomes of treatment. Our expert panel stipulated a sigmoidal relationship, estimating a very low success rate for durations < 4 months, approximately $25\text{--}40\%$ success at 1 year, depending on medication, and a maximum success rate of approximately $75\text{--}80\%$ by about 5 years. Consistent with these estimates and those of (70),

⁵ Specifically, (70) records that 28.8% of patients are successful, 36% drop out, and 18% are unsuccessfully transferred after median treatment durations of 39.5 weeks, 22.9 weeks, and 32.4 weeks respectively.

we parametrise the function for *Tx success fraction* (p_{\odot}^R) to match these estimates, with $m_R = 1$, $\lambda_R = 3.5$, $\kappa_R = 0.85$, and $p^{RM} = 0.8$ (see S2.d.ii.(3)). Note that this results in higher success rates for treatment durations < 4 months than the near-zero rates that our experts estimated; however, there is evidence that *some* success with treatment is possible even after short durations (71).

S3.c.iv.(2) Effects of treatment on mortality

We express the *effect of MOUD Tx on OD death rate / non-OD death rate* ($w_{\odot}^{TO} / w_{\odot}^{TN}$) as multipliers of the respective mortality rates for people with HUD or OUD not receiving MOUD treatment.

For non-OD mortality, we calculate an average aggregate non-OD mortality rate for untreated HUD and OUD groups of 1.43 people per 100 person-years. Based on reported hazard ratios, we calculate the effect of MOUD treatment (w_{\odot}^{TN}) at 0.54 for buprenorphine, 0.37 for MMT, and 0.93 for Vivitrol (see S3.e)).

For the effect on OD mortality (w_{\odot}^{TO}), several studies report no significant difference between hazard ratios for buprenorphine and methadone. As our calculated averages for each were very close (0.301 and 0.289, respectively), we instead use a combined average effect for buprenorphine and MMT of 0.295, with an effect of 0.439 for Vivitrol (see S3.e)). Note that for simplicity, we apply these multipliers to the overdose death rate, without modifying the base overdose rate. Insofar as the reductions reflect greater likelihood of resuscitation in the event of overdose, this is accurate; insofar as they may reflect reductions in baseline overdose rates, it likely means that non-fatal overdoses are being somewhat overestimated for people in MOUD treatment. This overestimation is, however, very small in absolute terms.

S3.c.v) Remission

Remission from disorder is a critical part of recovery, which ideally encompasses a return to functioning, health, and quality of life (72), though clinical remission is more narrowly defined as people who have had no symptoms of a substance use disorder for at least one year (1). The estimated millions of people who are in opioid use disorder remission (73) reflect the history of the crisis. They are both a potential role model and source of hope for others, and also remain at risk of relapse themselves.

We found no reliable time series data on size of populations in remission, and so remission is excluded from our panel of time-series data used in model estimation. Nevertheless, their inclusion in the model is important, and indeed model performance is improved when this group is retained rather than allowed to disappear from the system.

Initial values for remission stocks ($R_R, R_O, R_H, R_{RS}, R_{OS}, R_{HS}$) are estimated using findings from various papers analysing NESARC Waves I and II (see S3.e)). Once people enter remission, they are no longer distinguished by their treatment history. This distinction could be made in future iterations of the model, if data become available on rates of relapse after remission by treatment history and type.

Note that though remission does not require abstinence, NSDUH does not identify people who report non-disordered use who are in remission from use disorder. Due to this lack of data, and to reduce model complexity, we do not represent non-abstinent remission (which would entail non-disordered use) as a separate state, nor do we capture flows of people from UD states into non-disordered use states. As a result, NSDUH respondents who report non-disordered use while in remission will be counted in the corresponding non-disordered use population counts (M^y or N^y).

Remission via MOUD is a function of duration in treatment (see S2.d.ii.(3)). We specify the hazard rate of remission without MOUD for people with HUD (*remission rate HUD no MOUD Tx*, ρ_{URH}) at 0.068 per person-year, based on a systematic review and meta-analysis of SUD remission rates (72). We use their pooled result for conservative (low) estimates of remission rates, as their review encompasses all SUDs, but HUD tends to be more severe than other SUDs.

For people with Rx OUD, we estimate the hazard rate of remission (ρ_{URR}/ρ_{URO}) as part of the main model estimation process. We use the rate for people with HUD as a lower limit for this rate, to allow for the potential lower overall severity of Rx OUD compared to HUD.

After some time in the remission stocks, people transition to stable remission (R_{SR} , R_{SO} , R_{SH}), after which they are no longer at risk of relapse or overdose death. Evidence indicates that the risk of return to use disorder typically drops considerably after an average of at least five years in remission (74, 75), though with some variation and often taking longer. We therefore estimate the *time to stabilize remission* (τ_{RS}) as part of the main model estimation process.

S3.d) Overdoses, naloxone, and synthetics

S3.d.i) Overdose mortality data

We use annual multiple cause of death mortality data from CDC's National Vital Statistics System (NVSS) to estimate overdose death flows ($o_{(.)}$). The records in the NVSS microdata provide information on all deaths occurring within the United States, and each underlying cause of death is coded according to the International Classification of Diseases (ICD) classification system, Tenth Revision (ICD-10) (76).

We identified all drug-related fatal overdoses in NVSS mortality data using the following ICD-10 underlying cause of death codes: X40–X44, X60–X64, X85, or Y10–Y14. Among these records, we identified opioid-involved fatal overdoses by type[s] of opioid, using the following ICD-10 codes: prescription opioids or methadone (T40.2 or T40.3), heroin (T40.1), synthetic opioids other than methadone (T40.4), and unspecified opioids (T40.6) (77). We group these records into a set of mutually exclusive and collectively exhaustive combinatorial categories, which we then aggregate into four streams of annual deaths (see **Figure S11**) involving:

- 1) Prescription opioids or methadone, but *not* heroin or synthetics
- 2) Heroin but *not* synthetics, possibly involving prescription opioids or methadone
- 3) Synthetics but *not* heroin, possibly involving prescription opioids or methadone
- 4) Synthetics *and* heroin, possibly involving prescription opioids or methadone

Deaths involving only unspecified opioids are allocated in proportion to the size of these four categories each year. These four death data streams are read into the model to estimate overdose death flows.

Prescription opioids or methadone	Heroin	Synthetic opioids	Stream
			N/A
			1
			2
			3
			4

Figure S11: Allocation of overdose mortality by MECE categories to model data streams. Red indicates that a drug class is reported as involved in a death, and grey indicates it is not.

The NVSS data identify overdose deaths by the substance[s] involved, e.g., Rx opioids only, Rx opioids + heroin, heroin + synthetic opioids, synthetic opioids only, and so on. This creates a fundamental limitation – deaths are identified by the substance[s] involved in the last use episode[s] before death, not by the use *behaviour* that the decedent primarily engaged in, but SOURCE classifies people by use behaviour (e.g., Rx OUD vs. HUD). We allocate deaths involving a given substance to the user group[s] which primarily use that substance, recognising that this is a substantial simplifying assumption. People with Rx misuse or Rx OUD with or without heroin use (M , U_R , U_O) are assumed to contribute to Rx overdose deaths, while people with non-disordered heroin use or HUD (N , U_H) contribute to heroin deaths.

S3.d.ii) Synthetic death allocation structure

Synthetic-involved deaths present an additional challenge. We do not explicitly identify synthetic users. Most synthetic use, especially since ~2013, involves illicitly manufactured synthetics that have entered the drug supply, whether as an adulterant in or replacement for heroin, or possibly in the form of counterfeit prescription pills. Unfortunately, the CDC data (or any overdose death data to our knowledge) do not distinguish between prescription and illicitly manufactured synthetics.

Prior to ~2013, the vast majority of synthetic-involved deaths involved *only* synthetics, without other Rx opioids or heroin. All available evidence indicates that widespread fentanyl contamination of both pill and powder drug supplies only occurred after 2013 (33, 47, 48). We therefore assume that a small fraction of synthetic-involved deaths pre-2013 (specifically, those with co-reported heroin + synthetics) may have been due to low-level penetration of fentanyl in the illicit/powder drug supply (including a small but notable spike in 2005-2006 (78, 79)), but that the vast majority of synthetic-involved deaths at the time (i.e., those with no co-reported heroin) were due to intentional misuse of prescription fentanyl. This separation allows us to estimate the baseline rate of overdose due to prescription fentanyl (β_R^S), which we assume affects people with Rx OUD (80, 81).

Using β_R^S based on pre-2013 data, we can then separate synthetic-involved overdose deaths after 2013 into two streams:

- 1) a ‘base’ stream driven by intentional prescription synthetic use (without co-reported heroin), and
- 2) an ‘excess’ stream that combines
 - a. synthetic deaths without co-reported heroin, less the projected base stream, presumably driven by largely unintentional use of illicitly manufactured fentanyl, with
 - b. all synthetic deaths with co-reported heroin, driven by contamination of the heroin supply.

The latter two streams collectively account for the excess deaths attributable to fentanyl penetration through its effects on the process of overdose death (see S2.d.iii.(3)), which allows estimation of the effect sizes involved.

Note that we combine heroin deaths and excess synthetic deaths into a single stream for purposes of model estimation (see S4.b) for details).

S3.d.iii) Fentanyl penetration

We calculate a time series of fentanyl prevalence (*fentanyl penetration curve*, ϕ) using data from the National Forensic Laboratory Information System (NFLIS). NFLIS aggregates the number of reports of various drugs from forensic analyses of substances seized by law enforcement. We calculate ϕ as the

fraction of reports involving fentanyl or its analogues out of the total reports of heroin or fentanyl & analogues each year:

$$\phi = \frac{\text{Reports of fentanyl \& analogues}}{(\text{Reports of heroin} + \text{Reports of fentanyl \& analogues})} \quad (3.9)$$

Note that some portion of the reports of fentanyl & analogues may actually involve prescription fentanyl rather than illicitly manufactured fentanyl, as well as fentanyl pressed into counterfeit prescription pills as opposed to in powder form (see S2.d.iii.(3)). NFLIS data do not disambiguate reports by form or source, only substance. Because we cannot exclude these reports, ϕ is almost certainly an overestimate of powder-form, illicitly manufactured fentanyl as a fraction of heroin + fentanyl reports. However, several studies point to fentanyl exposure among heroin users being at least as great as indicated in NFLIS, if not much higher – at least 50% by 2017 (82–87). We therefore do not think the overestimation of ϕ due to prescription fentanyl or counterfeit prescription pills is of substantial concern.

S3.d.iv) Naloxone distribution

S3.d.iv.(1) Total kits distributed

We approximate total naloxone distributed using two data sources, corresponding to the two main channels by which naloxone kits enter the community – distribution through harm reduction and other community programs, and pharmacy purchases.

We calculate the former using the only published national data on naloxone kit distribution through community programs (88–90). These reports provide annual estimates of kits distributed for three years (2009, 2013, 2019). We extrapolate to other years from these data points using the annual percentage growth in programs and estimates reported in these three years. Note that after mid-2014, the only publicly available data are on *injectable* naloxone kits (i.e., not Narcan®) distributed by the OSNN naloxone buyer's club (90), a different sample of harm reduction programs than is reported on in 2012 and 2015.

For naloxone purchased in pharmacies, we use IQVIA NPA® data (see S3.b.i)) on prescriptions for naloxone filled in outpatient pharmacies (retail and mail-order).

These naloxone distribution totals do not include naloxone distributed to EMS, which are not reflected in sales data and for which, to our knowledge, time series data are not available at the national level. Naloxone is useful but not usually necessary for successful EMS intervention (91) (see S2.d.iii.(2)), so for simplicity, we do not include EMS naloxone use in the current version of the model.

S3.d.iv.(2) Naloxone kit allocation

Kits are not distributed equally between people who use prescription opioids vs. heroin (88), though we do not have a direct estimate of what fraction of kits go to heroin users (F^{NH}) vs. prescription opioid users, or more precisely, to people most likely to witness heroin user overdoses vs. prescription opioid user overdoses (e.g., including friends & family).

In order to estimate F^{NH} , we calculate the fraction of naloxone utilisation events involving heroin vs. prescription opioid overdoses, based on the total overdoses of each type multiplied by $p_W p_{IB(\cdot)}$. We anchor the estimate of F^{NH} using a prior value (see S4.a.ii)) for the fraction of utilisation events involving heroin overdoses. We calculate at this fraction at 86% based on the fraction of naloxone reversals reported to harm reduction programs involving heroin or something other than prescription opioids (88).

S3.d.iv.(3) Naloxone distribution efficiency

We derive the functional form for *probability Nx bystander...*, $p_{IB(.)}$ (see S2.d.iii.(2)) based on data from (92), which is the only estimate to our knowledge of how naloxone distribution affects probability of utilisation in the event of overdose. (92) reports the probability of naloxone utilisation in witnessed overdoses across 12 U.S. states as a function of naloxone kits distributed per 100,000 population. It also reports partial data on how probability of naloxone utilisation varies by distribution channel (standing order vs. prescription vs community distribution, though these data are insufficient to derive separate functions. An exponential function fits well with both the aggregated and disaggregated data reported.

Note that given the data available on total naloxone kits distributed (see above), this function results in naloxone kit utilisation fractions consistent with existing estimates from literature, which finds that 6-13% of all kits distributed are used (see S3.e)).

S3.d.v) Intervention Probabilities

We calculate *probability OD witnessed* (p_W) as the weighted average of the proportion of nonfatal overdoses that have been reported as witnessed (76.4%) across six studies (see S3.e)). Note that most of the data for this estimate come from older studies, as newer studies tend to report only the fraction of fatal overdoses that are witnessed. Conditioning on overdose fatality skews the reported probability compared to the unconditional probability (p_W), as whether an overdose is witnessed changes the net probability of death. Because the vast majority of overdoses are nonfatal, conditioning on overdoses being non-fatal results in less skew than conditioning on fatal overdoses, which are less likely to have been witnessed (41.1% weighted average; see S3.e)). In the absence of studies reporting aggregate witnessing probabilities for both fatal and nonfatal overdoses, we therefore use those with samples limited to the latter.

We calculate *probability of calling emergency services* (p_{IE}) in the event of a witnessed overdose using the weighted average from 31 studies that reported the fraction of all events witnessed during which the witness or someone else present called emergency services, yielding $p_{IE} = 42.4\%$ (see S3.e)). Most studies included people who use drugs, though some also included, e.g., friends and family members. Note that several studies were of people who had been trained in the use of naloxone, but at witnessed overdoses they reported only using naloxone and not also calling emergency services. Some evidence suggests possession of naloxone reduces the likelihood of calling emergency services (93), but we do not account for this potential interaction effect.

S3.e) Literature sources for parameter estimates

Various parameters in the model are synthesised from multiple studies following extensive literature searches. As a general rule, we sought to use multiple and diverse sources to compensate for the potential non-representativeness of study populations. Calculations and explanations for each parameter are described in the relevant sections above. Literature sources used are summarised in **Table S7**.

Table S7. Literature sources for parameter estimates

Parameter name	Parameter symbol	Parameter value	Sources
Average duration for buprenorphine	τ_{TB}	222.3 days, 0.61 years	(111, 112, 121–130, 113, 131, 132, 114–120)
Average duration for methadone	τ_{TM}	365 days, 1 year	(111, 116, 135–137, 118, 120, 125, 126, 129, 132–134)
Average duration for injectable naltrexone (Vivitrol)	τ_{TV}	82.4 days, 0.23 years	(115, 130, 138–141)
Average aggregate non-OD mortality rate for untreated OUD and HUD groups		1.43 per 100 person-years	(142–147)
Buprenorphine-waivered providers	B		(148–152)
Effect of MOUD treatment on non-OD mortality	w_B^{TN}	0.54	(143–147, 153–155)
Effect of MOUD treatment on non-OD mortality	w_M^{TN}	0.37	(143–147, 153–155)
Effect of MOUD treatment on non-OD mortality	w_V^{TN}	0.93	(143–147, 153–155)
Effect of MOUD treatment on OD mortality	w_B^{TO}, w_M^{TO}	0.295	(143–147, 153–155)
Effect of MOUD treatment on OD mortality	w_V^{TO}	0.439	(143–147, 153–155)
Initial value for remission stocks	$R_R, R_O, R_H, R_{RS}, R_{OS}, R_{HS}$		(156–160)
Naloxone kit utilization fraction		6–13%	(88, 161, 170, 171, 162–169)
Probability OD witnessed	p_W	76.4%	(172–177)
Probability fatal OD witnessed		41.1%	(178–182)
Probability of calling emergency services	p_{IE}	42.4%	(161, 163, 179, 183–191, 164, 192–201, 165, 202, 167, 169, 171, 175–177)

S3.f) Expert consultation process

We developed SOURCE’s structure through an extensive process of iterative consultation with subject-matter experts. In accordance with established best practices in model development (94), initial model development involved close engagement with client groups, consultations with subject-matter experts, and extensive review of existing literature, especially the few other models of the crisis extant at the time (95, 96). As part of this process, we interviewed 22 subject-matter experts within and outside the federal government, including clinicians, epidemiologists, and addiction and harm reduction experts. Initial development focused on identifying main feedbacks and areas of operational importance, while establishing a simulating (but not yet quantified) model structure. This stage of development lasted approximately nine months.

We then iteratively refined the model through 30 meetings with an expert working group that included five clinicians and addiction experts and three modelling consultants, as well as nine additional meetings with other individual subject-matter experts. These discussions addressed issues such as the causal structure of important phenomena, expert estimates of various parameters, and questions of data interpretation, as well as testing and evaluation of model estimates and behaviour. Along with primary quantification, this process of model refinement took approximately 15 months.

S4) Model Estimation

S4.a) Overview

SOURCE is a nonlinear and complex model, which makes finding an estimation framework with clean, closed-form analytical solutions highly challenging and unlikely. Instead, we estimate the model by maximum likelihood (97), using a Gaussian likelihood function to fit simulated time series to historical data, as well as a penalty term on a small number of point observations of certain key ratios.

The model can be thought of as a deterministic system of ordinary differential equations, with some set of unknown parameters (as well as known ones specified based on literature, expert estimates, etc., and exogenous time-series inputs). To avoid over-fitting, we do not use any time-varying parameter inputs (98), instead relying on endogenous feedbacks captured in the model structure to generate observed dynamics.

The maximum likelihood estimation framework identifies the most likely value for each unknown parameter, given the historical data. We combine this with a Markov Chain Monte Carlo (MCMC) simulation approach (99, 100) to identify the credible regions of parameter space and quantify uncertainties in parameter estimates and projections. We describe a synthetic data validation procedure aimed at building confidence in the estimation framework in S4.e).

S4.a.i) Likelihood function for historical data

The model generates simulated expected values for several time series variables, such as populations in several use states and various transition flows between states (see **Table S8** for full listing of time series used in estimation). Let μ_{it} represent simulated values for variable i at time t , while y_{it} represents the corresponding observed historical data points. With θ as the vector of unknown model parameters, we can summarise the model as a function f that yields predicted values for μ_{it} given θ as well as a set of exogenous time-series inputs x_{jt} (for variables j , e.g., prescribing rates, naloxone distribution; see **Table S8** and S3) for details):

$$\mu_{it} = f(\theta, x_{jt}) \quad (4.1)$$

We use a Gaussian (log-) likelihood function to specify the likelihood of observing y values given θ and x (which result in predictions μ):

$$LT(y_{it}|\theta, x_{it}) = \sum_{it} -\frac{(\mu_{it} - y_{it})^2}{2\omega_i^2} - \ln(\omega_i) \quad (4.2)$$

Summing the log-likelihood function over variables i and times t yields the full log-likelihood for observed data given a specific parameterisation and set of time-series inputs. The Gaussian function includes a set of scale parameters or calibration weights for each i variable (ω_i), which we approximate with the standard deviations of the corresponding observed time series y_{it} . These weight variables account for differences in the underlying magnitude and variability of the different time series used. Note that this estimation approach assumes i.i.d. error terms; it may therefore result in overly narrow credible intervals on the resultant parameter estimates (see S4.e)).

S4.a.ii) Penalty terms for key point observations

In addition to the likelihood value for observing historical data series, we also calculate a penalty term for certain key ratios and other point observations k (see **Table S9**). These penalty terms represent priors derived from literature estimates, expert judgment, or limited datasets, which we incorporate to provide some constraint on the estimation process:

$$LP(y_{kt}|\theta, x_{jt}) = - \sum_{kt} \frac{(\mu_{kt} - y_{kt})^2}{2\omega_k^2} \quad (4.3)$$

Here y_{kt} represents the prior expected values for k defined over specified time periods, which are compared against the model-generated values μ_{kt} . The scaling parameter ω_k represents the pre-specified allowable deviation of k values from their prior expected values.

While the selection of key observations k and their prior expected values y_{kt} and allowable variance ω_k are guided by existing data, there is inevitably an element of subjectivity in their selection. Such subjectivity, however, is not disqualifying. There is a degree of subjective judgment involved in any modelling endeavour, from problem definition to model specification to the estimation process, and indeed in all scientific endeavour in the first place. Incorporating priors in this manner allows us to inject valuable information into the estimation process without constraining it more than the quality of said information warrants. Absent the use of priors, we would either have to discard the informational value of these data points, or build them into hard constraints on parameters or fixed assumptions, neither of which seems like a desirable alternative. Instead, therefore, we aim to present the use of these assumed priors in a transparent manner while also validating the estimation procedure where possible (see S4.e)). Comparison of the model's simulated key observations k and their prior expected values y_{kt} are presented in S5.b), **Table S12**.

S4.b) Data used in estimation

Table S8 summarises the panel of time series data used in model estimation, whether as observed targets for model fitting y_{it} or as exogenous input variables x_{jt} . Sources and adjustments for these data are detailed in S3). The estimation period spans 1999-2020; see S6.d) for further information on the inclusion of the period of the COVID-19 pandemic.

Table S8. Panel of time-series data used in model estimation

Time series		Source	Model variable[s]
Observed data / calibration targets y_{it}	Rx misuse no PY heroin	NSDUH	M
	Nondisordered heroin use	NSDUH	N
	Rx OUD no PY heroin	NSDUH	$\sum_{(.)} U_R + (1 - F_{(.)}^R) T_{R(.)}, \quad (.) \in \{B, M, V\}$
	Rx OUD with PY heroin	NSDUH	$\sum_{(.)} U_O + (1 - F_{(.)}^R) T_{O(.)}, \quad (.) \in \{B, M, V\}$
	HUD	NSDUH	$\sum_{(.)} U_H + (1 - F_{(.)}^R) T_{H(.)}, \quad (.) \in \{B, M, V\}$
	Total buprenorphine patients	Various (see S3.c.i))	$\sum_{(.)} T_{(.)B}, \quad (.) \in \{R, O, H\}$
	Initiating Rx misuse own Rx	NSDUH	r_{MI}
	Initiating Rx misuse diverted	NSDUH	r_{MD}
	Total heroin initiation	NSDUH	$r_{ND} + r_{MN} + r_{UO}$
	Initiating heroin no Rx	NSDUH RDAS	r_{ND}
	Initiating heroin with Rx misuse	NSDUH RDAS	r_{MN}
	Initiating heroin with Rx OUD	NSDUH RDAS	r_{UO}
	Total overdose deaths base Rx	NVSS	$o_{mc} + o_M + o_{UR} + o_{UO} + o_{TR} + o_{TO}$
	Total overdose deaths synth base	NVSS	o_{URS}
	Total overdose deaths heroin & excess synthetics	NVSS	$o_{UN} + o_{UH} + o_{UNF} + o_{UHF} + o_{TH}$
	Total overdose deaths	NVSS	$\sum_{(.)} o_{(.)}$
Exogenous inputs x_{it}	Patients receiving opioid prescription	IQVIA	m_P
	Prescriptions per person	IQVIA	m_N
	Average days per prescription	IQVIA	m_D
	Average opioid MME per day	IQVIA	m_M
	ADF fraction of prescribed opioids	IQVIA	F^{AR}
	Buprenorphine-waivered treatment providers	Various (see S3.c.iii))	B
	Methadone maintenance treatment capacity*	N-SSATS	K_M
	Vivitrol® treatment capacity*	IQVIA	K_V
	Naloxone kits distributed	IQVIA, various	v_T
	Heroin price index (1999 = 1)	UNODC, STRIDE	$1/D_{AH}$
	Fentanyl penetration	NFLIS	ϕ

Table S9. Point data used as priors y_{kt} in model estimation

Prior	Model variable[s]	Year	Value	Source
Nonfatal OD ratio Rx	$\sum_{(.)} \frac{\eta_{R(.)}}{\omega_{R(.)}}$	< 2013	35	(203, 204)
Nonfatal OD ratio heroin	$\sum_{(.)} \frac{\eta_{H(.)}}{\omega_{H(.)}}$	< 2013	30	(203, 204)
Bup demand fulfilment ratio	$\frac{r_{UTB}}{r_{UTB}^*}$	2018	0.587	(67)
Probability Nx bystander heroin	p_{IBH}	2019	0.2	Expert judgment
Nx utilization events H user fraction	$\frac{\sum_{(.)} (n + o)_{H(.)} p_W p_{IBH}}{p_W \sum_{(.)} ((n + o)_{H(.)} p_{IBH} + (n + o)_{R(.)} p_{IBR})}$	2013	0.86	(88)
Rx OUD in remission total	$R_R + R_O + \sum_{(.)} F_{(.)}^R (T_{R(.)} + T_{O(.)}) ,$ $(.) \in \{B, M, V\}$	2013	893153	(159, 205)
HUD in remission total	$R_H + \sum_{(.)} F_{(.)}^R T_{H(.)} ,$ $(.) \in \{B, M, V\}$	2013	284174	(159, 160)
Rx OUD in stable remission total	$R_{SR} + R_{SO}$	2013	1349830	(159, 205)
HUD in stable remission total	R_{SH}	2013	485323	(159, 160)

S4.c) Iterative estimation procedure

With 53 estimated parameters and an additional 20 initial stock corrections, the model is sizeable, but not so large as to make searching the full parameter space computationally impractical. Nonetheless, to speed up the estimation process, we use a multi-step iterative procedure (**Figure S12**), estimating partial models first (101) in order to converge on the most likely region of parameter space before estimating the full parameter vector simultaneously. The goal of this process is first to identify the location in parameter space of the global peak in the likelihood surface, i.e., the most likely parameter set given the data and model structure, and second to define the credible region or hypervolume in parameter space around that maximum-likelihood peak. All steps prior to 7) use the Powell direction search method implemented in Vensim™ simulation software.

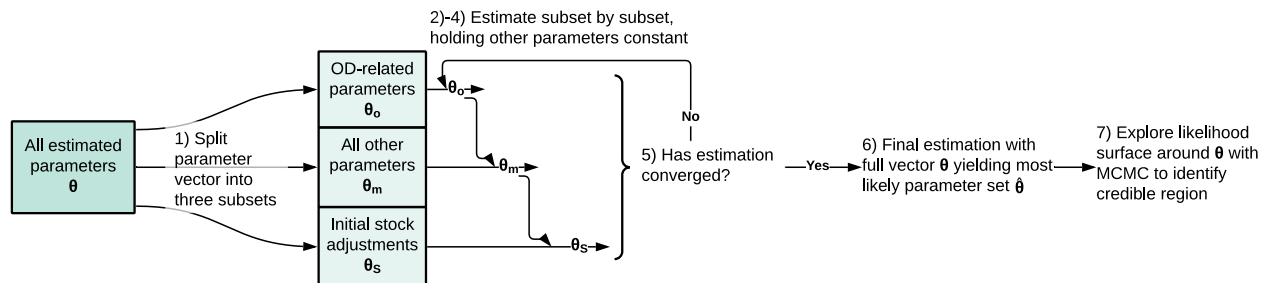


Figure S12. Overview of iterative estimation procedure

- 1) We split the parameter vector θ and the set of data variables used in calibration y_{it} into a few subsets. Specifically, we define θ_o as the subset of parameters that directly affect overdose death

risks (as explained in S2.d.iii)), with y_{ot} being the data variables directly tracking overdose deaths; we also define θ_S as the subset of *initial stock correction* parameters ($m_{0(\cdot)}$), explained further below. Remaining parameters (excluding θ_o and θ_S) we define as θ_m .

- 2) We first estimate only the subset of parameters θ_o that directly affect overdoses, to maximise the likelihood of observing y_{ot} , holding all other parameters constant at their previous best-estimate values. The first time we perform this step during each estimation process, overdose deaths are calculated using exogenous data values $y_{(\cdot)t}$ for all drug user stocks; subsequently, we use endogenously generated stock values. This is in effect a partial model calibration aimed at matching just the overdose death data using only overdose-related parameters. With 11 overdose-related parameters included in this step out of 53 estimated parameters total, it helps to narrow down the plausible range of parameter space.
- 3) Next we estimate only θ_m , using the full set of target data y_{it} , holding θ_o constant at the values estimated in step (2).
- 4) Next we estimate only the initial stock correction parameters θ_S , using the full set of target data y_{it} , holding θ_o and θ_m constant at previously estimated values. The initial stock corrections modify the baseline initial stock values ($S_{(\cdot)0}^*$), which are derived from values of data at the initialisation of the run in 1999:

$$S_{(\cdot)0} = m_{0(\cdot)} S_{(\cdot)0}^* \quad (4.4)$$

$$S_{(\cdot)0}^* = y_{(\cdot)0} \quad (4.5)$$

The correction parameters are necessary to allow initial stock values to differ from the first observed data points. These first data points ($y_{(\cdot)0}$) are not inherently any more accurate than any other values for these stocks observed in the data ($y_{(\cdot)t}$), and are equally subject to random variation like process noise and measurement error. The estimation process accounts for such randomness over time; however, using the first data points directly as the initial stock values would effectively over-weight those first points, asserting that the random variation contained in their values is of zero magnitude. The initial stock corrections ($m_{0(\cdot)}$) provide a means to avoid this problem, giving the first data points the same importance as any others.

- 5) We iterate through steps (2)-(4), each time holding constant the parameter subsets (θ_o , θ_S , θ_m) not being estimated in the current step at their last-estimated values, until the iterations cease to offer significant improvement (approx. 0.05% of total log-likelihood) when compared at step (3). At that point, we assume the estimation has converged to close to the optimal parameter set, speeding the subsequent steps. Step (4) is repeated one more time before moving on.
- 6) We conduct a full optimization using the complete parameter vector θ and comparing the full set of time-series data y_{it} , starting from the parameter values estimated in the last iteration in step (5). This full optimization locates the exact peak in the full likelihood landscape, which corresponds to the best-fit maximum likelihood parameter set $\hat{\theta}$ for the full model.
- 7) Finally, we carry out an MCMC simulation to explore the likelihood surface in parameter space around $\hat{\theta}$. We use an MCMC algorithm designed for exploring high-dimensional parameter spaces using differential evolution with self-adaptive randomised subspace sampling (99). We use an extensive burn-in period of 1500000 MCMC samples, by which point the MCMC chains yield stable outcomes (Gelman-Rubin PSRF < 1.1 for 95.0% and < 1.2 for 98.8% of chains) (102); we then continue the MCMC to sample a further 1000000 outcomes, and then randomly take a subsample

of 5000 of those 1000000 sampled points to use for sensitivity analyses and projections (see S6)). We also use this subsample to derive 90% credible intervals for parameter estimates.

This process is automated using a Python script that controls the simulation software (Vensim). We conduct the analysis using a parallel computing feature of Vensim on a multi-core Windows machine. Full analysis code is available online at <https://github.com/FDA/SOURCE>.

S4.d) Quantifying uncertainties

S4.d.i) Credible intervals for parameter estimates

Estimated parameters can have interacting effects on the total likelihood and overall fit of the model to observed data. Maximum likelihood parameter estimates are therefore not independent, but should be thought of as a parameter set $\hat{\theta}$. Similarly, the MCMC simulation explores the high-likelihood credible *region* of parameter space, producing a sample or subsample of credible parameter *sets*. This credible region provides a more meaningful quantification of uncertainty in parameter estimates than univariate ranges, which are in effect the projections of the credible region onto each parameter's axis. We therefore utilise the MCMC subsample of credible parameter sets for projections and sensitivity analyses, detailed below.

However, reporting a high-dimensional credible region is impractical and difficult to present meaningfully. For transparency, we report here the univariate 90% credible intervals for each parameter (**Table S10** in S5.a)), with the caveat that there may be substantial covariation between different estimated parameters (due to, e.g., compensatory effects) that the univariate intervals will miss. The full MCMC subsample that defines this region is available online at <https://github.com/FDA/SOURCE>.

S4.d.ii) Estimating measurement error

The estimation procedure identifies the region of parameter space that results in the highest likelihood of observing the data y_{it} given the model f . The model, however, is deterministic, and does not account for random process noise nor measurement error. Model-generated predictions based on the maximum-likelihood parameter set, $\hat{\mu}_{it}$, therefore represent expected values for observed variables y_{it} , rather than predictions or projections of the exact unobserved realisations of variables i , which will include process noise, or of observed variables y_{it} , which will include measurement error as well. Similarly, projections based on the credible region of parameter space around $\hat{\theta}$ capture uncertainty in the expected values of observed variables, rather than the full range of uncertainty in possible trajectories for those variables, which includes the aforementioned sources of randomness.

In order to make projections that better express the range of possible trajectories, therefore, we need to account for such randomness. We do this by injecting random noise into model projections for variables i , with a unique realisation of this noise stream for each parameter set in the credible region used in projections or sensitivity analyses.

To parametrise the distribution of this noise term, we fit a multivariate Normal distribution to the residuals from the main model estimation process. Use of a multivariate rather than independent univariate Normal distributions is important as many of the observed variables draw on the same few data sources (e.g., NSDUH, NVSS), so there is likely to be substantial covariance in their measurement errors. On the other hand, while autocorrelation is a common issue with time-series data, the long interval

between data points (1 year) relative to the speed of underlying processes means there is little autocorrelation in the residuals, with most sources of inertia in the data accounted for by model mechanisms. We therefore use noise terms without autocorrelation.

For a handful of estimated variables, there is a clear systematic bias in the temporal pattern of the residuals (see S5.b)). Reasons for these biases are largely understood and excluded from current model scope, as discussed further in S5.b). In order to accurately approximate the random, unbiased component of measurement error (and generate realistic synthetic data; see below), before fitting the noise distribution, we first detrend the residuals using a fitted 2nd-order polynomial function. We then fit the multivariate Normal distribution to the detrended residuals.

S4.e) Synthetic data validation

To build confidence in our estimation framework, we conducted a synthetic data experiment to better assess its accuracy. We generate synthetic data representing artificial ‘parallel universes’ by simulating the model with known parameter values combined with simulated measurement noise. We then attempt to recover those parameter values from the simulated data using our exact estimation framework. We can then assess how well the estimated parameters and credible intervals correspond to the known, true values.

S4.e.i) Data generation and estimation

To generate the synthetic data, we first randomly draw 20 parameter sets θ^s from the MCMC subsample generated in the main model estimation process as described in S4.b). Since these parameters sets are drawn from the credible region of parameter space, they provide plausible alternatives similar but not identical to the model’s estimated most-likely parameter set $\hat{\theta}$. We then simulate the model using these parameter sets, injecting ‘measurement’ noise into simulated model outputs to create realistic ‘observed’ values for data, y_{it}^s . The noise stream for each synthetic data set is randomly drawn from the same multivariate Normal distribution estimated for the residuals from the main model estimation process, as described in S4.d.ii).

The synthetic data sets thus generated are available online at <https://github.com/FDA/SOURCE>.

We then estimate the model using each synthetic data set y_{it}^s in turn, in place of observed data. We use the same set of exogenous time-series inputs x_{jt} for each estimation. Estimation follows the same procedure as for the main model, described in S4.b). As with the main estimation, we start with uninformed priors on all parameters (uniform distributions with large ranges).

S4.e.ii) Synthetic data estimation results

The synthetic data estimation process covers a total of 1460 parameters (20 synthetic data sets x (53 parameters + 20 initials each)). In most cases, the estimation process recovers results close to the ‘true’ parameter values from the synthetic data, with the median distance between estimated and true values expressed as a percentage of the width of the estimated 95% credible interval at 28%.

Overall, estimated credible intervals contain true values at close to, but slightly under, the theoretically expected rates – the 50%, 80%, 90%, 95%, and 98% CIs contain true parameter values 35%, 60%, 72%, 78%, and 82% of the time respectively (**Figure S13**). The under-fitting indicates that estimated CIs are on the whole slightly narrower than they should be, i.e., that the estimation of CIs is slightly over-confident.

The most likely driver of this overconfidence is the existence of covariance between different time series used in estimation, in both errors and expected values. The error covariance is captured in the measurement noise estimation (see S4.d.ii)) and hence the synthetic data. The estimation process, however, treats each time series used as an independent component of the likelihood function. As such, the estimation process ‘overstates’ the informational content of the time series data used in estimation (which is lowered due to covariance), resulting in overconfident, i.e., overly narrow credible intervals. In addition, it is possible that there is some over-dispersion of error terms, such that an alternative likelihood function (e.g., exponential, negative binomial) may be more suitable for estimation and would yield more theoretically accurate (i.e., wider) CIs.

In light of these factors, while not perfect, the estimation process performs very well – in benchmark tests on a far simpler SEIR model, similar Gaussian maximum likelihood estimates performed far worse (<20% true values in 50% CI, and <40% in 90% CI) (103). This result therefore gives some confidence that our estimated credible intervals, while on the narrow side, are reasonable approximations of the most likely ranges of true parameter values.

Complete results of the synthetic data estimation exercise are presented in S5.d).

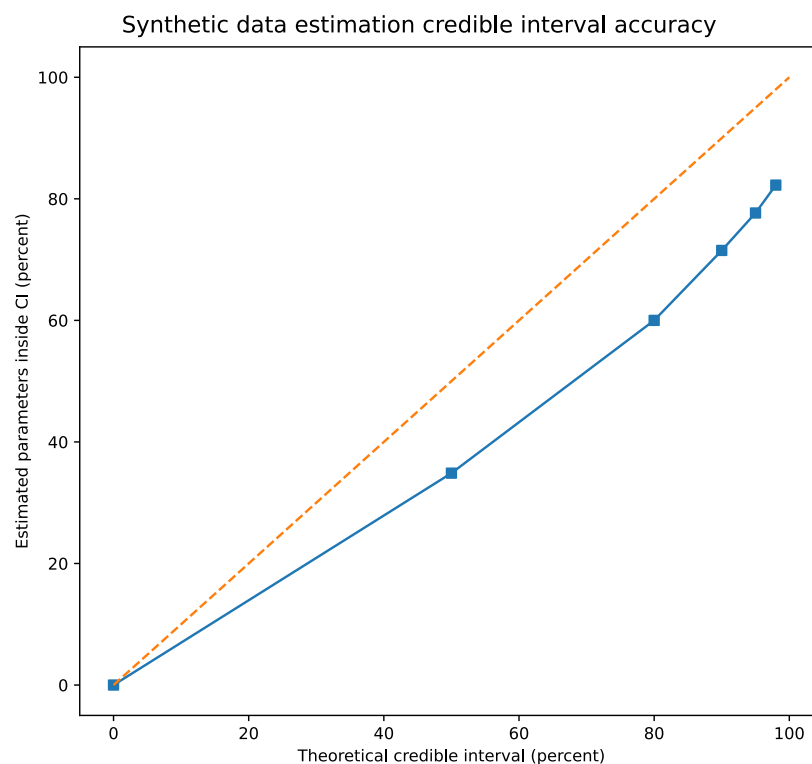


Figure S13. Theoretical vs. actual percentage of parameters contained within different credible intervals. The dashed 1:1 diagonal would indicate perfect matching between theoretical and estimated CIs.

S4.f) Out-of-sample validation

To build further confidence in our model estimates and projections, we also conducted an out-of-sample test using a holdout dataset to assess the model’s predictive capabilities.

This test should be approached advisedly. Feedback-driven dynamic models like SOURCE seek to endogenously capture the shifting drivers of important system processes over time. However, even with as broad-boundary a model as SOURCE, there remain several major drivers of the crisis which the model treats exogenously, such as MOUD capacity or, most importantly, fentanyl prevalence. Where emergent features of the opioid crisis arise primarily from these exogenous changes (rather than endogenous dynamics), the model will have little basis for quantitatively predicting the impacts. In addition, the temporal sparseness of the relevant data – most of which are only available annually – can limit the precision of certain parameter estimates. Nevertheless, the model’s ability (or failure) to predict observed trends remains informative.

S4.f.i) Estimation & assessment process

The first step of the out-of-sample test is estimating model parameters and uncertainties exactly as described in S4.a)-S4.d) above, but using calibration data y_{it} only from 1999-2012 (inclusive), using the remainder of the data as the holdout dataset for comparison.

We then use the MCMC subsample of credible parameter sets and estimated measurement noise (see S4.d)) to project a set of credible ranges for predicted observed data y_{it}^p for 2012-2020. This process is similar to synthetic data generation (see S4.e.i)), but uses the full MCMC subsample rather than just 20 individual parameter sets. Note that as with synthetic data generation, we use the actual exogenous input variables x_{jt} to generate these predicted values.

Finally, we compare the predicted observations and credible ranges y_{it}^p to actual observed data y_{it} in the holdout dataset, to assess the model’s performance.

S4.f.ii) Out-of-sample validation results

The holdout dataset contains a total of 120 data points (15 observed time series⁶ over 8 years each). Of these, 72% (86/120) fall within the predicted 95% credible intervals.

While the level of predictive accuracy could be higher, it is reasonably good considering that the estimation period up to 2012 excludes virtually the entirety of the surge in illicit fentanyl, which started in ~2013. As such this predictive performance is despite the estimation having little basis for accurate estimation of fentanyl’s lethality and identification of relevant parameters.

More importantly, the model correctly projects trend changes in several variables, including Rx opioid overdose deaths, OUD / HUD, and heroin initiation from prior Rx use (**Figure S19**). The ability to predict shifts in trends not yet evident in the estimation data is a good indicator that the model’s structure accurately reflects important components of the actual system structure (104), i.e., the data-generating processes. Such evidence of structural soundness in turn gives confidence in the model’s ability to project future developments.

Complete results of the out-of-sample validation exercise are presented in S5.e).

⁶ The model estimation uses 16 observed time series altogether (see **Table S8**), but one series, *total overdose deaths synth base* (o_{URS}), ends in 2012, and so is absent from the holdout dataset.

S5) Full Results

S5.a) Full parameter estimates

Table S10 shows most likely values $\hat{\theta}$ as well as medians and 90% credible intervals for all 53 estimated parameters and 20 initial stock value adjustments θ . In some cases, the most likely value falls outside the 90% credible interval. While uncommon, this is not inherently erroneous – it could for instance indicate a ‘cliff-shaped’ likelihood surface, shallow-sloped on one side of its highest point and dropping off steeply on the other.

Table S10. Complete list of estimated parameter values & credible intervals

	Value	0.05	0.5	0.95
ADF effect strength initiating heroin with Rx OUD	1.00E-06	1.00E-06	0.1357	0.5355
Base survival probability H OD relative to Rx	0.9806	0.9625	0.9833	0.9923
Base survival probability Rx OD	0.9734	0.9670	0.9734	0.9776
Developing HUD rate no Rx OUD	0.05	0.05	0.0532	0.0646
Developing HUD rate with Rx OUD	0.5345	0.4413	0.5403	0.6466
Developing Rx OUD rate	0.0297	0.0251	0.0284	0.0322
Fentanyl effect on base survival max relative to H	0.7358	0.6054	0.7579	0.7923
Fentanyl effect on OD rate H max	2	2	2.0179	2.0941
Fraction Nx kits to H users	0.9371	0.9109	0.9337	0.95
Heroin availability strength developing HUD	1.00E-06	1.00E-06	0.0325	0.1581
Heroin availability strength initiating NDHU no Rx	1.00E-06	1.00E-06	0.0246	0.1064
Heroin availability strength net quit NDHU	1.4641	1.1020	1.5075	1.9014
Initial stock correction[RXM]	1.1476	1.0613	1.1567	1.2
Initial stock correction[NDH]	1.2	1.1344	1.1850	1.2
Initial stock correction[OUB]	1.1326	0.8124	1.0361	1.2
Initial stock correction[OUM]	0.8	0.8	0.9892	1.1908
Initial stock correction[OUV]	1.0603	0.8036	1.0068	1.2
Initial stock correction[OUT]	0.9841	0.8638	0.9803	1.1091
Initial stock correction[OUR]	0.8	0.8	0.8225	0.9272
Initial stock correction[OUS]	0.8	0.8	0.8110	0.8479
Initial stock correction[OHB]	0.9197	0.8	0.9423	1.1746
Initial stock correction[OHM]	0.8	0.8	0.9779	1.1880
Initial stock correction[OHV]	1.0532	0.8	0.9746	1.1919
Initial stock correction[OHT]	1.2	0.8237	1.0459	1.2
Initial stock correction[OHR]	0.8	0.8	0.9053	1.1513
Initial stock correction[OHS]	0.8	0.8	0.9196	1.1669
Initial stock correction[HUB]	0.9486	0.8	0.9675	1.1890
Initial stock correction[HUM]	1.2	1.0890	1.1760	1.2
Initial stock correction[HUV]	1.0960	0.8	0.9631	1.1918
Initial stock correction[HUT]	0.8	0.8	0.8071	0.8369
Initial stock correction[HUR]	1.2	0.9183	1.1256	1.2
Initial stock correction[HUS]	1.2	1.1379	1.1863	1.2
Initiating heroin no Rx base	119578	113597	119109	124251
Initiating Rx misuse diverted base	1879840	1800000	1870839	2040037
Initiation rate heroin with Rx misuse	0.0134	0.0121	0.0130	0.0142
Initiation rate heroin with Rx OUD relative to Rx misuse	4.2679	3.8660	4.3846	4.9072
Initiation rate Rx misuse own Rx	0.0346	0.0297	0.0347	0.0405

Net quit rate heroin with Rx misuse	0.0924	0.0469	0.1020	0.1532
Net quit rate heroin with Rx OUD	0.1023	0.0334	0.0932	0.1660
Net quit rate NDHU	0.5	0.4461	0.4899	0.5
Net quit rate Rx misuse	0.1432	0.1192	0.1454	0.1742
NSDUH misuse redefinition fixed effect	0.4401	0.3940	0.4442	0.4930
Nx kit distribution efficiency	0.0003	0.0003	0.0003	0.0004
Overdose rate base HUD	0.1674	0.1152	0.1801	0.1996
Overdose rate NDHU relative to HUD	0.25	0.25	0.2592	0.3082
Overdose rate base Rx misuse	0.001	0.001	0.0015	0.0034
Overdose rate base Rx OUD	0.2513	0.1985	0.2492	0.2976
Overdose rate synth baseline	0.0072	0.0051	0.0078	0.0092
Perceived risk strength initiating heroin with Rx use	0.9277	0.8017	0.9296	1.0431
Perceived risk strength initiating NDHU no Rx	0.2714	0.1893	0.2933	0.4201
Perceived risk strength initiating Rx misuse diverted	0.5556	0.4323	0.5313	0.6678
Perceived risk strength initiating Rx misuse own Rx	0.5659	0.3599	0.5898	0.8234
Perceived risk strength net quit heroin with Rx OUD	1.00E-06	1.00E-06	0.1725	0.4665
Perceived risk strength net quit NDHU	1.00E-06	1.00E-06	0.0083	0.0392
Perceived risk strength net quit NDHU with Rx	1.00E-06	1.00E-06	0.0159	0.0829
Perceived risk strength net quit Rx misuse	0.5592	0.3590	0.5607	0.7615
Relapse rate HUD	0.5830	0.5054	0.5678	0.6410
Relapse rate Rx OUD relative to HUD	0.1375	0.1247	0.1868	0.3595
Remission rate Rx OUD relative to HUD	1	1	1.1434	1.5645
Rx availability strength developing Rx OUD	1.2152	0.9569	1.2665	1.6151
Rx availability strength initiating Rx misuse	0.3883	0.0154	0.3385	0.7456
Rx availability strength net quit Rx misuse	1.0443	0.6509	1.0552	1.4777
Rx vs H availability strength developing HUD with Rx OUD	1.2648	0.7996	1.2011	1.6651
Rx vs H availability strength initiating heroin with Rx OUD	1.00E-06	1.00E-06	0.0169	0.0805
Rx vs H availability strength initiating NDHU with Rx	0.1890	0.0346	0.1997	0.3738
Social influence strength developing HUD	1.00E-06	1.00E-06	0.0146	0.0629
Social influence strength developing Rx OUD	1.00E-06	1.00E-06	0.0317	0.1387
Social influence strength initiating heroin with Rx OUD	1.9518	1.7373	1.9203	2
Social influence strength initiating NDHU no Rx	0.1208	0.0061	0.1370	0.3317
Social influence strength initiating NDHU with Rx	1.7896	1.5926	1.8060	1.9780
Social influence strength initiating Rx misuse	0.7902	0.4611	0.7923	1.1234
Time to stabilize remission	9.7425	8.8793	9.6251	10.4528
Tx seeking rate Rx OUD no H total	0.4932	0.3754	0.5194	0.8204

S5.b) Fit to historical data

Figure S14 shows fit between simulated model output μ_{it} and historical data y_{it} for all time-series data used in model estimation (see Table S8 in S4.b)), spanning 1999-2020.

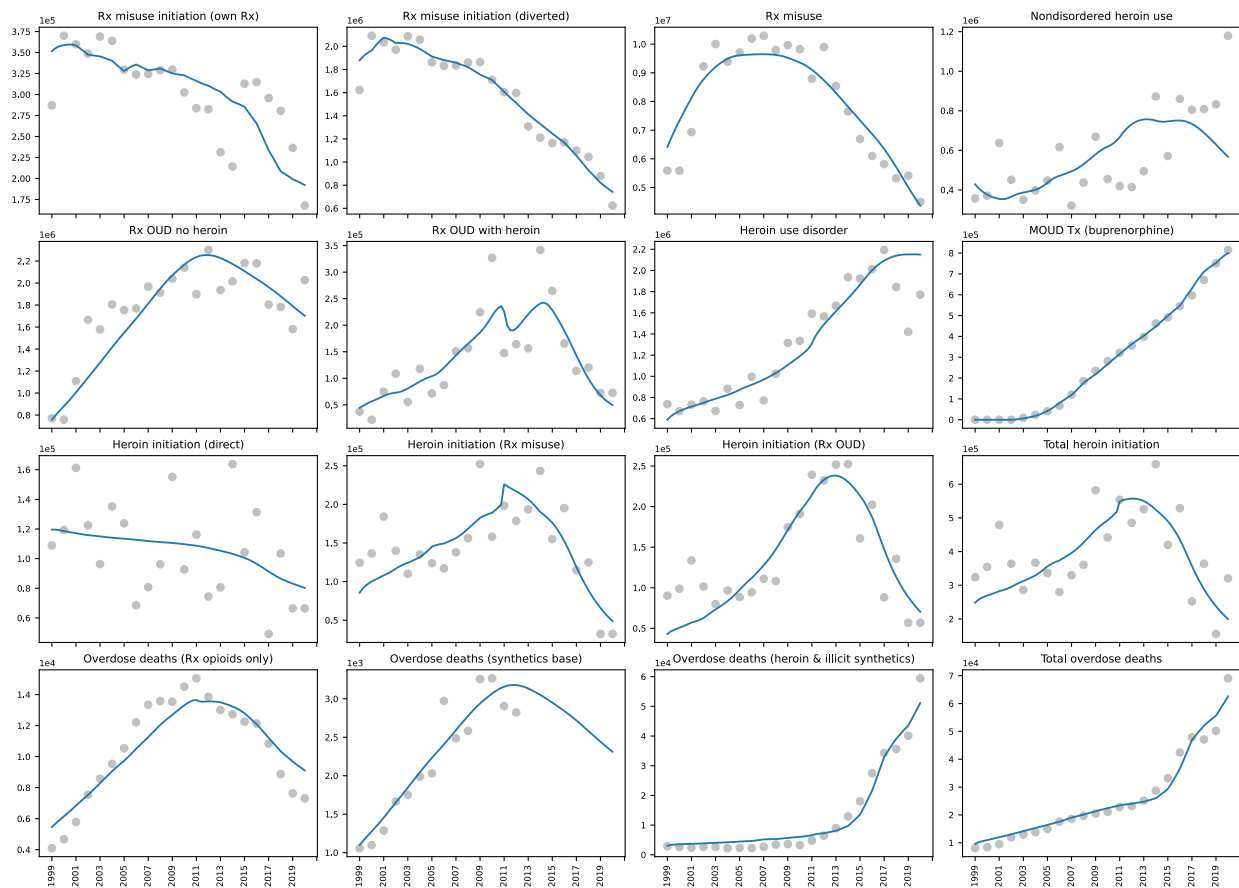


Figure S14. Comparison of simulated model output (blue) to historical data (grey) for all time-series data used in estimation

Table S11 reports full quality-of-fit metrics for each estimated time series. Most of the time series fit well, with the majority of error stemming from unequal covariance (U_c), indicating an unbiased estimation (21, 105).

In a few cases there is substantial unequal variance (U_s) or bias (U_m). Notably, for overdose flows, there is some degree of systematic skew. The most likely reason for this is cohort effects on the likelihood of overdose which we do not incorporate in our model, particularly for people with Rx OUD. For Rx-involved overdose deaths, trends in the prescribing of benzodiazepines and their joint use with prescription opioids may play some role as well. For heroin & illicit synthetic overdoses, anecdotal reports indicate changes in the purity and potency of heroin in the years around 2010, when mortality data show sharper rises in overdose deaths than our estimates. However, reliable data on heroin purity are extremely difficult to obtain, especially at regional level, so we have been unable to test for this effect.⁷

⁷ We examined several other potential explanations, including the roles of drug availability, prevalence of injection drug use, prevalence of polysubstance use, and a ‘naivete’ effect with new initiates being more vulnerable to overdose; none of these were able to account for the skew, or more precisely, the divergence between trends in use disorder prevalence and overdose mortality, which are generally otherwise closely correlated.

Table S11. Goodness of fit statistics for each estimated time series

	MAEN	MAPE	R2	MSE	Um	Us	Uc
Rx misuse	0.070	0.077	0.896	4.70E+11	0.005	0.288	0.707
Rx OUD no heroin	0.099	0.102	0.775	4.81E+10	0.035	0.050	0.915
Rx OUD with heroin	0.254	0.310	0.718	2.03E+09	0.002	0.170	0.828
Nondisordered heroin use	0.266	0.269	0.201	4.14E+10	0.005	0.145	0.850
Heroin use disorder	0.116	0.119	0.836	4.89E+10	0.007	0.032	0.960
MOUD Tx (buprenorphine)	0.031	0.112	0.998	1.81E+08	0.007	0.072	0.921
Rx misuse initiation (own Rx)	0.098	0.111	0.506	1.51E+09	0.001	0.000	0.999
Rx misuse initiation (diverted)	0.047	0.055	0.953	8.58E+09	0.008	0.012	0.980
Total heroin initiation	0.193	0.211	0.513	7.67E+09	0.030	0.012	0.958
Heroin initiation (direct)	0.223	0.249	0.192	7.99E+08	0.001	0.500	0.499
Heroin initiation (Rx misuse)	0.205	0.242	0.582	1.25E+09	0.015	0.022	0.963
Heroin initiation (Rx OUD)	0.189	0.241	0.754	1.08E+09	0.015	0.004	0.981
Overdose deaths (Rx)	0.094	0.114	0.894	1.39E+06	0.000	0.373	0.627
Overdose deaths (heroin & illicit synthetics)	0.193	0.424	0.969	9.28E+06	0.008	0.221	0.771
Overdose deaths (synthetics base)	0.088	0.086	0.900	5.80E+04	0.027	0.056	0.917
Total overdose deaths	0.083	0.094	0.969	8.19E+06	0.010	0.101	0.889

Table S12 reports simulated values for key point observations used as prior expected values in the estimation process (see S4.a.ii)). Most of the simulated values are close to their prior expected values in the relevant years, with the exception of Rx OUD in remission, which is somewhat lower than expected.

Table S12. Simulated values for key point observations

Prior	Year	Prior Value	Simulated Value (5-95 CI)
Nonfatal OD ratio Rx	< 2013	35	36 (29-43)
Nonfatal OD ratio heroin	< 2013	30	31 (22-38)
Bup demand fulfilment ratio	2018	0.587	0.587 (0.350-0.765)
Probability Nx bystander heroin	2019	0.2	0.20 (0.19-0.21)
Nx utilization events H user fraction	2013	0.86	0.86 (0.83-0.88)
Rx OUD in remission total	2013	893153	853840 (809155-903527)
HUD in remission total	2013	284174	348994 (333881-377700)
Rx OUD in stable remission total	2013	1349830	1436970 (1423930-1516760)
HUD in stable remission total	2013	485323	415899 (395961-440437)

S5.c) Base case projections

Figure S15 shows base case projections under ‘exogenous trends continue’ (ETC) assumptions for all time-series data used in model estimation (see **Table S8** in S4.b)). As noted in the main text, these projections should not be considered a precise forecast. Note that the projected trajectories and credible intervals reflect projected *expected values* for observed variables, with uncertainty arising only due to uncertainty in parameter estimates, rather than exact observed / reported realizations of those variables, which will also include uncertainty due to measurement noise; see S4.d.ii) for further details.

The projections show an overall decline in opioid use, use disorder, and mortality for both prescription and illicit opioids. This decline arises in significant part because of the continued effects of perceived risk and social influence driving falls in heroin use that appear to already be under way. This projected pattern

depends in part on the reliability of the most recent 1-2 years of heroin use data from NSDUH (see S3.a.iv.(1)), which should temper interpretation of these trends. However, even with some rebound or increase in use in 2020 presumably due to the effects of the COVID-19 pandemic, the projected decline nonetheless persists.

The ETC projections include a slight rebound in prescription opioid use later this decade. Under ETC assumptions, levels of prescribing will fall somewhat from current levels, but remain significant (**Table S13**). This scenario thus presumes continued exposure of large numbers of people to potentially highly addictive prescription opioids, which combined with an attenuating risk response, may lead to an eventual rebound in prescription opioid misuse.

In addition to ETC projections, we also test an ‘optimistic’ and ‘pessimistic’ input assumptions scenario, shown in **Figure S16** and **Figure S17**. The ‘optimistic’ scenario assumes greater increases in MOUD capacity and naloxone availability, larger reductions in opioid analgesic prescribing, and a smaller increase in fentanyl prevalence compared to ETC assumptions, and vice-versa for the ‘pessimistic’ scenario (see **Table S13** for full details). The resultant differences in projected trends are instructive.

Table S13. Exogenous inputs with alternative base case assumptions for projections

Exogenous Input Variable	Source	2020 value	2032 Assumed Value***		
			ETC	Optimistic	Pessimistic
Fentanyl penetration	NFLIS	56.2%	80.7%	69.8%	99.5%
Naloxone kits distributed	IQVIA, various*	2.30 million	3.60 million	4.22 million	2.94 million
Heroin price index (1999 = 1)	UNODC, STRIDE	0.49	0.49	0.58	0.40
Buprenorphine-waivered treatment providers	Various*	94,200	178,300	224,900	134,500
Methadone maintenance treatment capacity**	N-SSATS	360,000	646,000	765,000	528,000
Vivitrol® treatment capacity**	IQVIA	32,900	45,800	52,700	39,900
Patients receiving opioid analgesic prescription	IQVIA	41.3 million	28.4 million	22.3 million	35.1 million
Prescriptions per person	IQVIA	3.49	3.31	3.01	3.50
Average days per prescription	IQVIA	24.4	26.8	24.0	28.0
Average opioid MME per day	IQVIA	31.3	23.6	20.2	28.0
ADF fraction of prescribed opioids (% of MME)	IQVIA	4.9%	3.1%	3.1%	3.1%
<p>* See S3.c.iii) and S3.d.iv.(1) for full details of data sources & calculations</p> <p>** Neither MMT nor Vivitrol capacity data are directly available; instead we calculate capacity based on treatment utilization data from the sources listed; see S3.c.iii)</p> <p>*** Broadly, the ‘optimistic’ scenario assumes stronger trends (1.5x ETC) in naloxone distribution, MOUD treatment capacity, and downward-trending aspects of prescribing, and weaker trends (0.5x ETC) in fentanyl penetration and upward-trending aspects of prescribing; vice-versa for the ‘pessimistic’ scenario.</p>					

Notably, in the ‘pessimistic’ case, while overdose deaths involving heroin and illicit synthetics rise substantially over the next few years, they nonetheless peak and start to decline. The increase is driven almost entirely by rapidly increasing fentanyl prevalence (see also S6.c)) in this scenario leading to growing overdose hazard and lethality for people who use heroin. Eventually, though, the falling prevalence of

heroin use (due to declining initiation) outpaces the increasing overdose hazard, and mortality starts to decline. For this reason, even if ongoing effects of the pandemic lead to increases in use, use disorder, or mortality through 2021, the projected declines are likely to occur (albeit from a higher peak) within the next few years.

The three scenarios also show noticeable differences in the prevalence of prescription opioid misuse and use disorder, as well as prescription-opioid-driven overdose deaths. These differences indicate that future trends in prescribing could substantially affect the trajectory of opioid use, with the potential for a more rapid rebound if the decade-long trend of falling opioid prescribing were to be reversed. Even with these rebounds, however, overall opioid-involved overdose mortality still declines even in the ‘pessimistic’ case despite rising prescription-opioid-driven deaths.

We must reiterate that despite the robustness of the projected declines in overdose mortality to different input scenarios, these projections categorically must not be interpreted as downplaying the severity of the crisis or lessening the need for intervention, as they still entail hundreds of thousands of deaths over the next decade. Instead, they highlight that the goal of policy should not merely be to achieve declines in mortality, but to do so faster and sooner than would otherwise occur anyway.

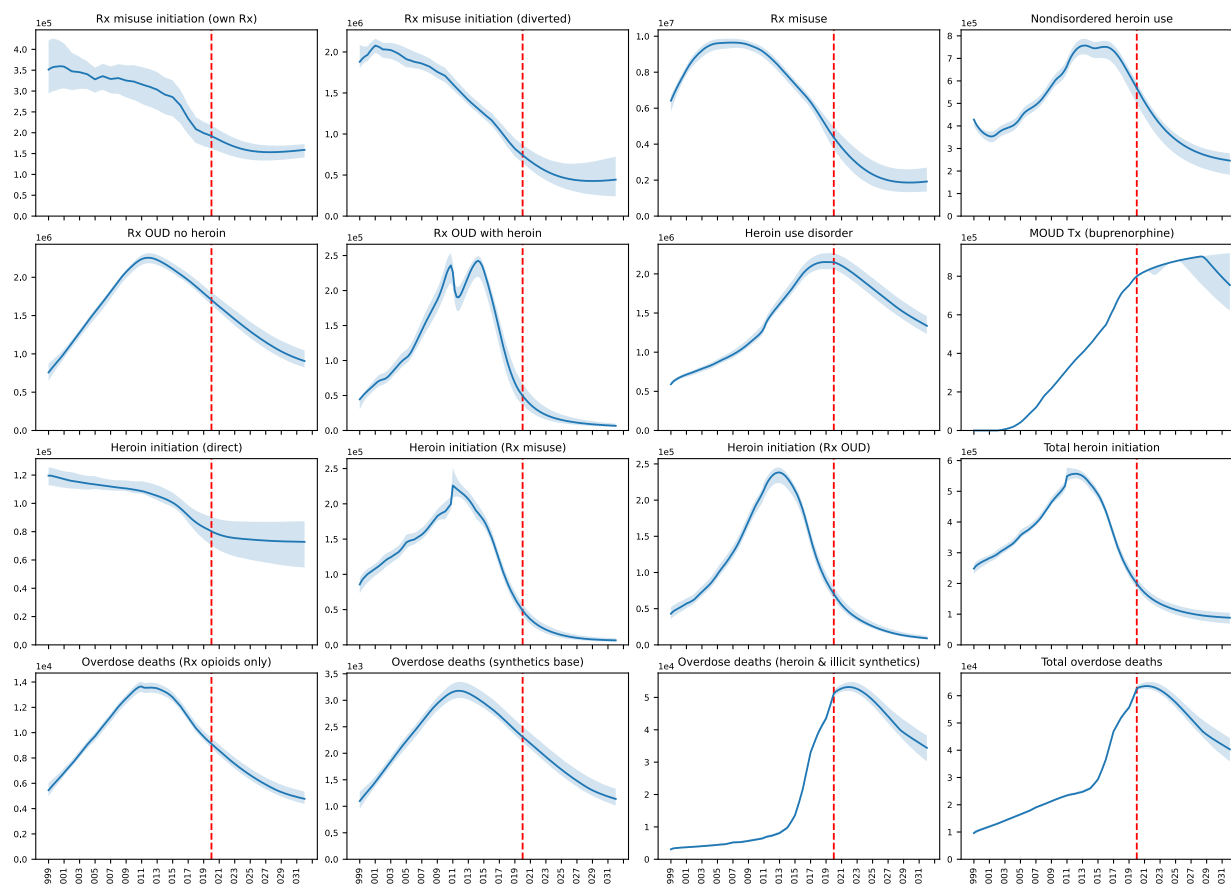


Figure S15. Base case ('exogenous trends continue') projections for all time-series data used in estimation. Shaded areas indicate 95% credible intervals, not accounting for measurement noise.

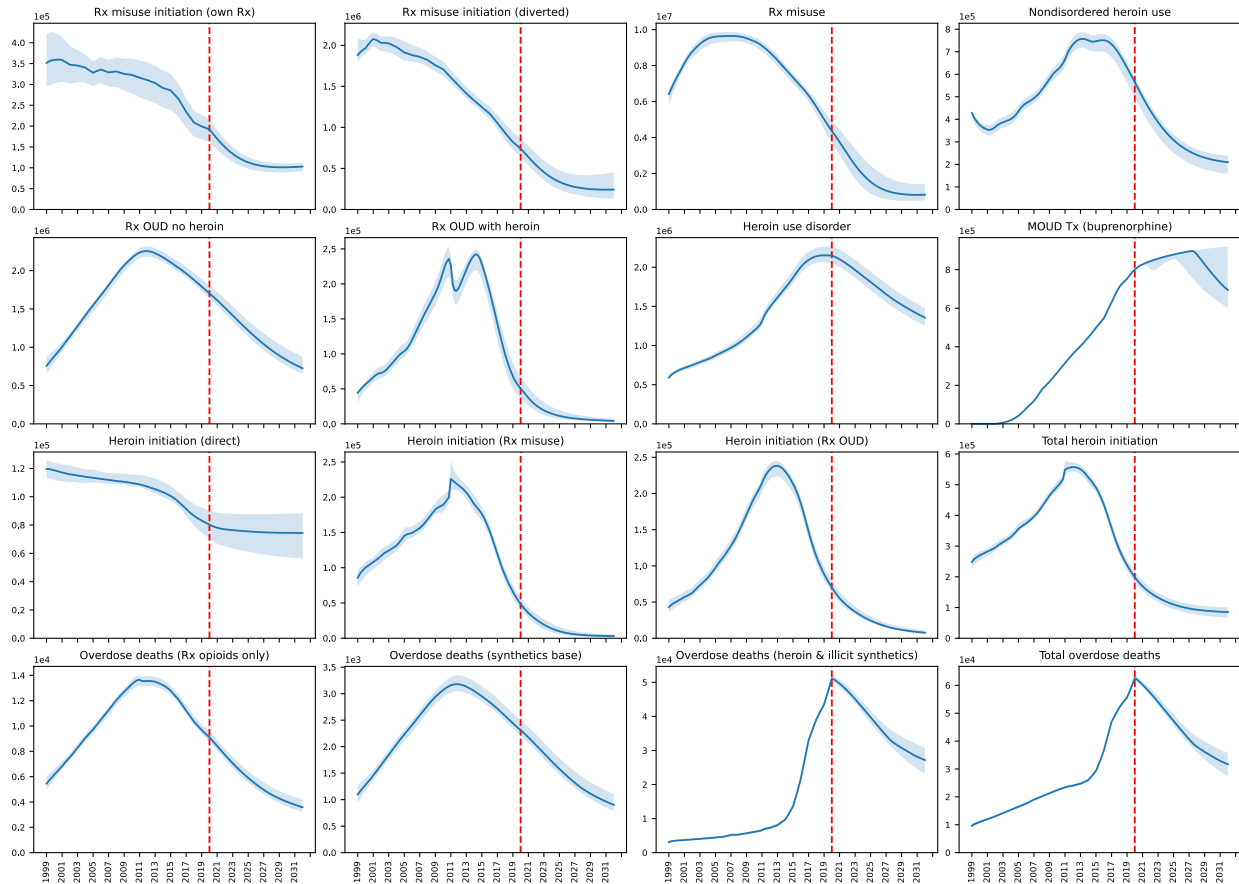


Figure S16. 'Optimistic' case projections for all time-series data used in estimation. The 'optimistic' scenario assumes more rapid reductions in opioid prescribing, more rapid increases in MOUD capacity and naloxone availability, less rapid growth in IMF prevalence, and reductions in heroin availability. Shaded areas indicate 95% credible intervals, not accounting for measurement noise.

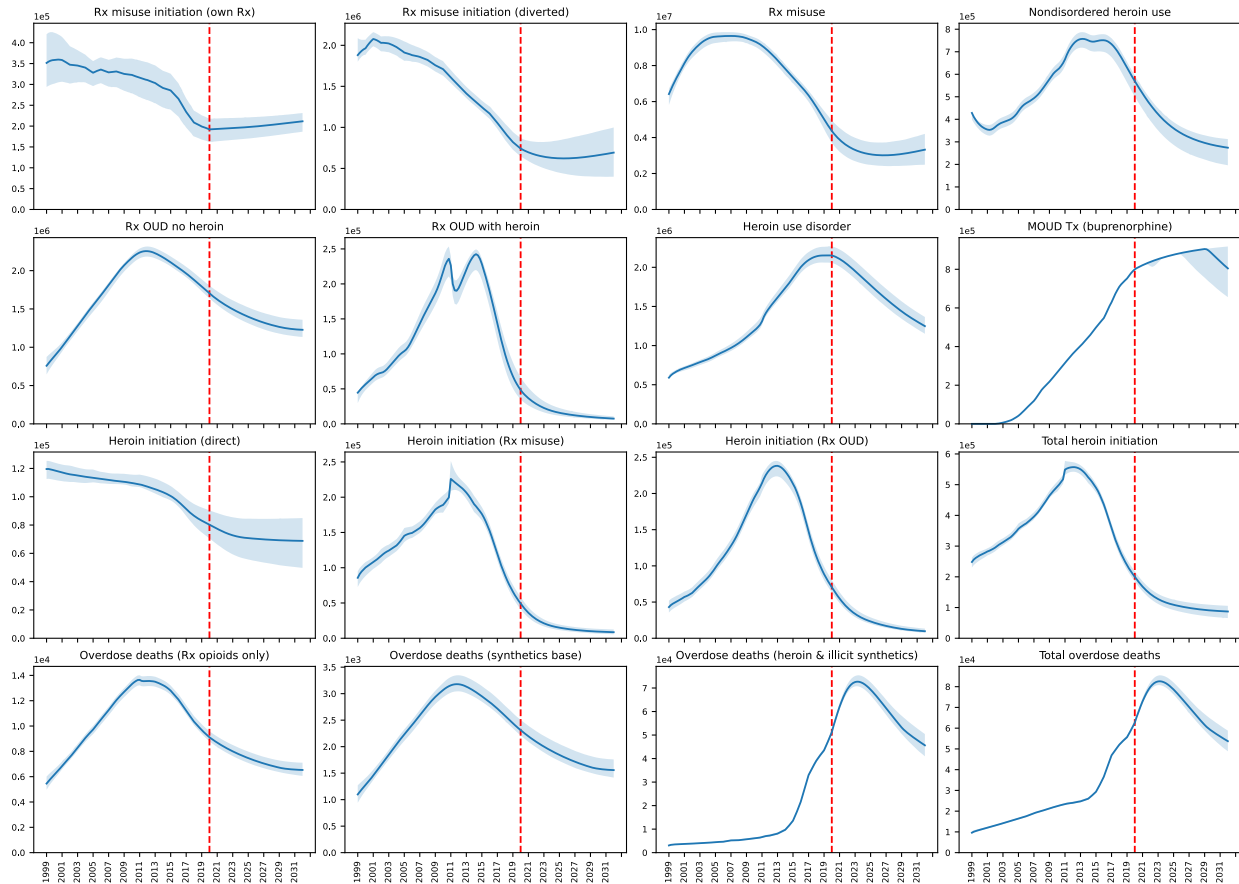
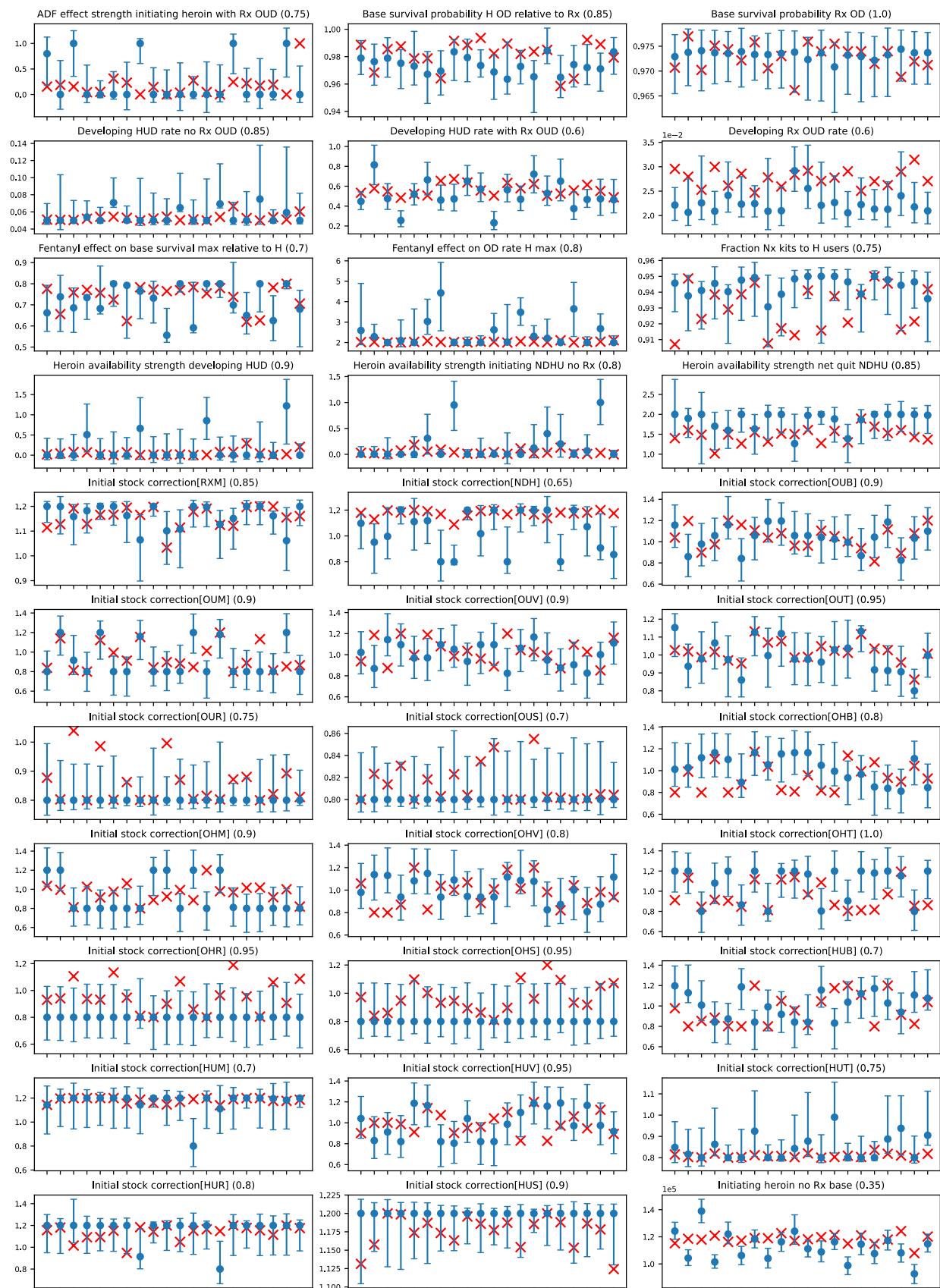
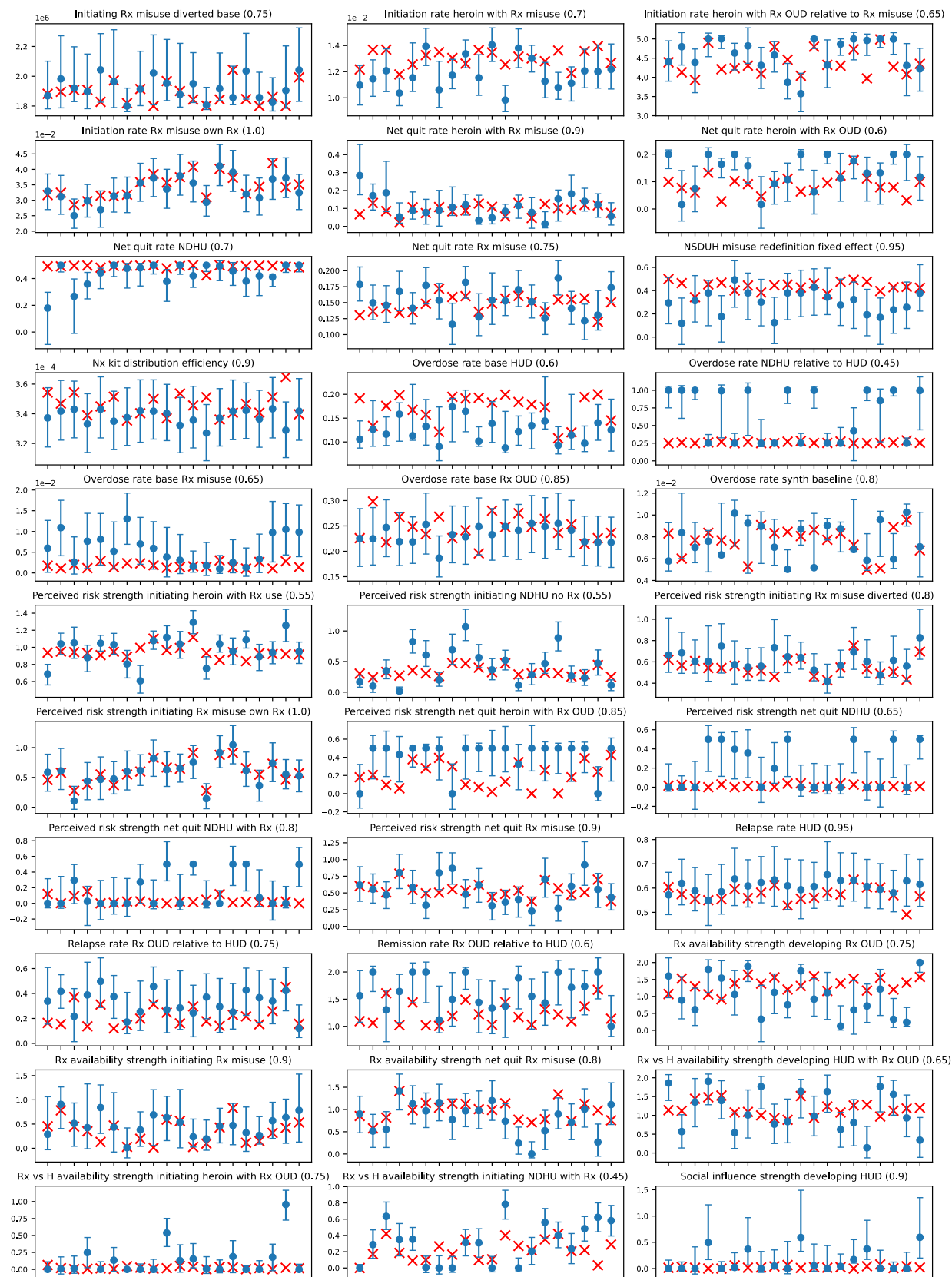


Figure S17. 'Pessimistic' case projections for all time-series data used in estimation. The 'pessimistic' scenario assumes growth in opioid prescribing, slower increases in MOUD capacity and naloxone availability, more rapid growth in IMF prevalence, and increases in heroin availability. Shaded areas indicate 95% credible intervals, not accounting for measurement noise.

S5.d) Full synthetic data estimation results

Figure S18 shows complete results of the synthetic data estimation exercise (see S4.e)) for all 20 synthetic data sets and all estimated parameters. Most estimated parameters values are close to their 'true' synthetic values across the board, with their 95% CIs containing 'true' values in most of the 20 runs. Several of the parameters that perform poorly in this regard are feedback effect strengths with 'true' values close to zero (e.g., some availability and social influence strengths), which due to their exponential formulation (see S2.c)) have little effect at such values. Relapse and remission rates, which have little data to anchor them in the estimation and can compensate for each other to some degree, also perform relatively poorly.





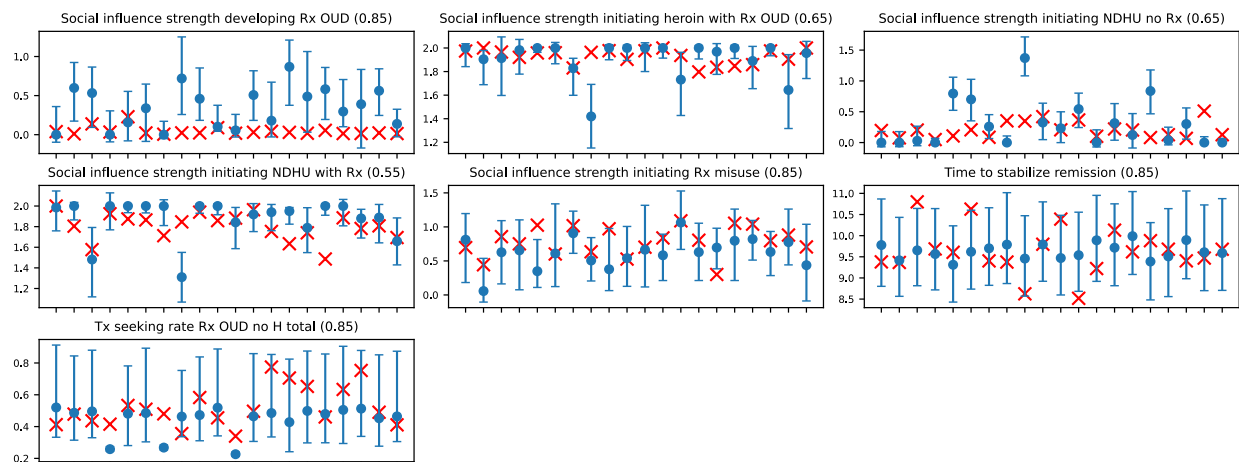


Figure S18. Estimated values & 95% credible intervals for all synthetic data parameters compared with ‘true’ values (estimated value & CI = blue dots & bars; true value = red crosses). Numbers in parentheses after each variable name indicate the fraction of true values for that parameter falling within estimated 95% CIs.

S5.e) Full out-of-sample validation results

Figure S19 shows complete results of the out-of-sample validation test (see S4.f)), comparing all model output μ_{it} and predicted 95% credible intervals against historical data y_{it} for all time-series data used in model estimation. Note that the figure shows credible intervals for estimated underlying or *expected* values $\hat{\mu}_{it}$ prior to 2012, and predicted *observed* values y_{it}^p for 2012-2020.

As discussed in S4.f.ii), most data points in the holdout dataset fall within the predicted 95% credible intervals. Furthermore, many of the missed datapoints fall just outside those intervals, e.g., for Rx opioid overdose deaths, heroin initiation from prior Rx misuse / Rx OUD / in total, and heroin & illicit synthetic / total overdose deaths for the period from ~2013-2016. The model correctly projects trend changes in Rx OUD (with and without heroin), HUD, heroin initiation from prior Rx misuse / Rx OUD / in total, and Rx opioid overdose deaths.

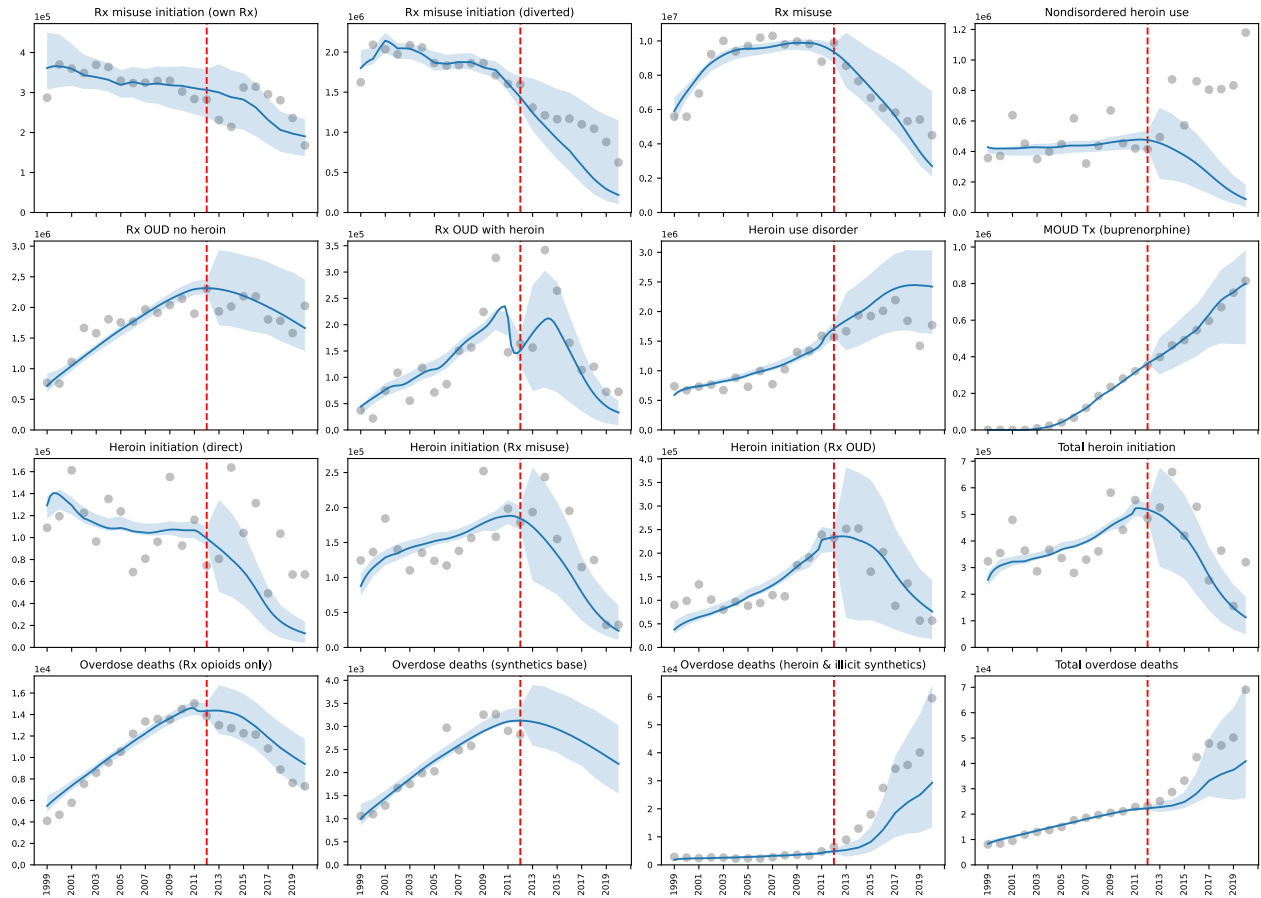


Figure S19. Comparison of model projections (blue) against historical data (grey), showing predictive accuracy for out-of-sample validation test. Bands are 95% Crls for estimated underlying values (estimation portion, before 2012) and for projected reported data (after 2012); projected reported values account for measurement noise, which is present in the actual holdout data.

S6) Sensitivity Analyses

S6.a) Sensitivity of outcomes to parametric assumptions

SOURCE includes 42 parametric assumptions drawn from literature and/or expert input (see S3). To assess the sensitivity of the model to these assumptions, we conducted sensitivity analyses on several of these parameters, varying them by $\pm 10\%$ and re-estimating the model in accordance with S4.c).

To keep the analysis computationally tractable, we do not test all 42 parametric assumptions, instead focusing on those with major structural roles or which may drive nonlinear changes.

We exclude several parametric assumptions which have conceptually important policy implications but whose specific values have little impact on the rest of the model. For instance, the *Tx seeking barrier loss fraction* (F^L) and its components are of great practical importance for increasing access to treatment, but changes in the assumed value of F^L will simply result in a proportional change in the estimated value of the *Tx seeking rate* ($\rho_{T(\cdot)}$) with no further impact on the model (see S2.d.ii.(1)):

$$r_{UT(\cdot)}^* = \rho_{T(\cdot)}(1 - F^L)U_{(\cdot)} \quad (2.13)$$

As such, we exclude it and other structurally similar parameters from the parametric sensitivity analysis.

We report the impacts of parametric changes on

- 1) Four main substantive outcomes:
 - a. Estimated cumulative overdose deaths and cumulative person-years of use disorder from 1999-2020
 - b. Projected cumulative overdose deaths and cumulative person-years of use disorder from 2020-2032, under 'ETC' case assumptions (see S6.c))
- 2) Median and maximum relative changes in estimated parameter values across all estimated parameters, excluding feedback strength parameters estimated at near-zero values (see below)
- 3) Average and maximum MAEN of all estimated historical time series, to identify impacts on overall goodness-of-fit as well as any outsize impacts on particular time series (see S5.b))

We report the first two sets of these impacts (1) and (2) as elasticities, i.e., fractional changes in outcome measure divided by fractional change ($\pm 10\%$) in the input parametric assumption. This yields a dimensionless value with a useful heuristic interpretation – absolute values < 1 indicate modest sensitivity, whereas absolute values > 1 indicate greater sensitivity meriting closer inspection.

Note that some feedback strengths ($\varepsilon_{(\cdot)}$) in the main model (see S2.c)) are estimated at near-zero values, which indicates that feedback has no significant effect (see S2.c.iv)). In these cases, re-estimation can yield seemingly large relative changes in $\varepsilon_{(\cdot)}$ which are actually insignificant in practice:

$$(D_{C(\cdot)})^{1E-06} \approx (D_{C(\cdot)})^{1E-05} \approx (D_{C(\cdot)})^{1E-04} \approx 1 \quad (6.1)$$

As such, in order to avoid skewing the reported elasticities of estimated parameters with apparent multiple-order-of-magnitude changes that are actually insignificant, we exclude from the average and maximum parameter elasticity calculation any parameters with baseline estimated value $< 1E-04$. This threshold restricts the exclusion to only feedback strength parameters $\varepsilon_{(\cdot)}$.

S6.a.i) Parametric sensitivity results

Table S14 reports elasticities of major outcomes and parameter estimates to changes in parametric assumptions, as well as changes in model goodness-of-fit (MAEN).

Overall model goodness-of-fit is virtually identical regardless of small changes in parametric assumptions. Similarly, overdose deaths and use disorder prevalence over the historical period of 1999-2020 are essentially unchanged. Taken together, these two results indicate the model is largely insensitive to precise values of parametric assumptions. Changes in those precise values can be easily accommodated with small changes in the values of estimated parameters, as indicated by the low median elasticity of parameter estimates.

The absolute maximum elasticity across all estimated parameters is distinctly higher. Despite that, the elasticity of key projected outcomes from 2020-2032 remains generally low, with absolute magnitudes well below 1 in most cases. The absolute elasticities for projected outcomes are two orders of magnitude larger than those for historical outcomes, but this is to be expected – firstly, the projected values are not directly constrained by historical data, and secondly, projecting forward amplifies the effect of minor divergences in parameter estimates.

Only in one case is there substantial sensitivity (elasticity ~ 1) to a parametric assumption – increasing the assumed duration of buprenorphine treatment results in an almost proportional increase in projected deaths.

This result may seem counterintuitive, as increased duration (and hence efficacy) of treatment might be expected to result in *fewer* deaths. To understand this result, it is useful to think of model estimation as a balancing or goal-seeking process, with historical data as the goal. If we assume historically greater treatment efficacy, the estimation process compensates by estimating more severe rates of use disorder development, relapse, etc. to maintain historically observed levels of use disorder. Those more severe or stronger counterbalancing forces, projected out into the future, result in *more* deaths than in the base case.

Notwithstanding that, on the whole, we find that SOURCE is largely insensitive to minor changes in parametric assumptions, which should help allay any concerns about any imprecision in the calculation of those values.

Table S14. Sensitivity of key outcomes to $\pm 10\%$ changes in parametric assumptions. Reported elasticities are averages of absolute value change with increase/decrease in parameter; polarity of change assumes 10% increase from assumed value.

	Cumulative overdose deaths	Cumulative UD person-years	Projected cumulative overdose deaths	Projected cumulative UD person-years	Med elasticity	Max elasticity	Avg MAEN	Max MAEN
ADF substitutability factor	-0.009	-0.006	-0.004	-0.005	0.048	-2.569	0.127	0.266
Effect of MOUD Tx on OD death rate[Bup]	0.001	-0.004	-0.034	0.008	0.032	-1.534	0.127	0.266
Effect of MOUD Tx on OD death rate[MMT]	-0.001	-0.001	-0.019	0.008	0.025	-1.988	0.127	0.266
Effect of MOUD Tx on OD death rate[Viv]	0.001	-0.003	-0.014	0.016	0.067	-2.629	0.127	0.266
OxyContin withdrawal magnitude	-0.002	0.009	-0.016	-0.011	0.037	2.985	0.127	0.266
Perceived risk decrease time	-0.002	-0.006	-0.007	-0.006	0.041	-2.040	0.127	0.266
Perceived risk increase time	-0.003	0.009	0.014	0.005	0.031	-0.800	0.127	0.266
Perceived risk weight NFOD	-0.002	-0.004	-0.008	-0.015	0.045	-2.227	0.127	0.266
Probability OD witnessed	0.002	-0.012	0.072	-0.027	0.040	-1.605	0.127	0.267
Probability of calling emergency services	-0.001	-0.002	-0.012	-0.008	0.035	2.159	0.127	0.266
Remission rate HUD no MOUD Tx	-0.002	-0.009	-0.090	-0.135	0.038	-2.207	0.127	0.266
Rx demand HUD with Rx OUD or misuse	-0.002	-0.004	0.013	0.033	0.020	-1.962	0.127	0.266
Rx demand Rx OUD no H	-0.001	-0.006	-0.010	0.023	0.025	-1.724	0.127	0.266
Sensitivity of Rx supply to patients receiving prescription	-0.002	-0.005	0.038	0.079	0.043	-2.294	0.127	0.266
Sensitivity of Rx supply to Rx per person	-0.001	-0.008	-0.018	-0.034	0.045	-1.952	0.127	0.266
Sensitivity of Rx supply to days per prescription	0.002	-0.003	0.017	0.026	0.029	-2.203	0.127	0.266
Sensitivity of Rx supply to MME per day	-0.001	-0.003	-0.034	-0.096	0.047	-2.493	0.127	0.266
Time to readjust Rx street supply	-0.005	0.004	-0.009	0.007	0.049	2.436	0.127	0.266
Tx average duration Bup	0.003	-0.002	-0.031	-0.050	0.048	-1.939	0.127	0.266
Tx average duration MMT	-0.001	-0.006	-0.005	0.013	0.045	2.157	0.127	0.266
Tx average duration Viv	0.002	-0.007	-0.015	-0.008	0.031	-1.850	0.127	0.266
Tx seeking fraction Bup HUD	0.000	-0.006	-0.010	-0.028	0.025	-1.887	0.127	0.266
Tx seeking fraction Bup Rx OUD	0.003	0.003	-0.011	0.037	0.053	-1.791	0.127	0.266
Tx seeking fraction MMT HUD relative	-0.001	-0.009	0.010	0.007	0.052	-1.482	0.127	0.266
Tx seeking fraction MMT Rx OUD relative	-0.003	-0.003	-0.012	-0.002	0.025	-1.627	0.127	0.266
Tx seeking rate HUD relative to Rx OUD no H	-0.002	-0.001	-0.019	-0.020	0.056	1.941	0.127	0.266
Tx success fraction inflection	-0.002	-0.006	-0.035	-0.163	0.070	-3.156	0.127	0.266

S6.b) Feedback loop knockout analyses

SOURCE's feedback structure is central to its behaviour, projections, and insights about historical trajectories. With multiple feedbacks acting simultaneously, sometimes with similar results, it can be difficult to disambiguate the impacts of different feedback processes through model estimation alone; some form of loop dominance analysis is warranted (106). To assess the relative role of different feedbacks in SOURCE and build confidence in our estimates of their strengths, we conducted loop knockout analyses (21, 106), examining how model behaviour and estimates change in the absence of certain feedbacks.

We test SOURCE's feedbacks in three groups, corresponding to the three main conceptual sets of feedback processes – social influence, perceived risk, and availability (see S2.c)). We deactivate all loops in one group at a time.

For each of these three sets of loops, we conduct two forms of loop knockout test: 1) running the model with baseline parameter estimates, with the focal loops deactivated; and 2) deactivating the focal loops and re-estimating the model in accordance with S4.c). The first test helps highlight the role played by the focal loops in the baseline model, while the second test highlights the ability of the other feedbacks in the model to compensate for the missing focal feedback, potentially acting as 'shadow' feedback structures (106).

For each test, we report the changes in the outcomes outlined in S6.a) above: 1) substantive outcomes for 1999-2020 and 2020-2032; 2) changes in estimated parameter values, excluding near-zero values (note that for the first test described, parameter values will not change); and 3) MAEN across historical time series. Unlike in S6.a), as there is no meaningful input change against which to normalise an elasticity value, we report the first two sets of impacts as fractional changes rather than elasticities.

S6.b.i) Loop knockout results

Table S15 presents summary outcomes of loop knockout analyses, while **Figure S20** shows model fit to data.

In all three 'deactivated' tests, model fit worsened considerably and both historical and projected key outcomes changed substantially, indicating that all three loop sets play important roles in generating the model's behaviour. These changes are greatest when perceived risk feedbacks are deactivated, consistent with recent heroin use data indicating a very strong risk response being vital in shaping the trajectory of the crisis (see S3.a.iv.(1)). Deactivating availability effects reduces both historical and projected deaths and use disorder, affirming the idea that the ready supply of both prescription opioids and heroin has been partly responsible for the scale of the crisis. Interestingly, deactivating social influence feedbacks has relatively little overall impact on cumulative deaths or use disorder, reflecting the fact (or more accurately, the structural assumption) that social influence goes both ways, and has contributed to both increases and decreases in drug use over the last ~20 years.

As expected, model fit in the 'recalibrated' tests worsens compared to baseline (baseline average MAEN = 12.7%), but only slightly; it worsens slightly more for the perceived risk feedback knockout than the others. Similarly, most estimated parameters do not change very much. Together, these results indicate some ability of model feedbacks to compensate for each other, as discussed above. Closer examination of how exactly the feedbacks change, however, is instructive.

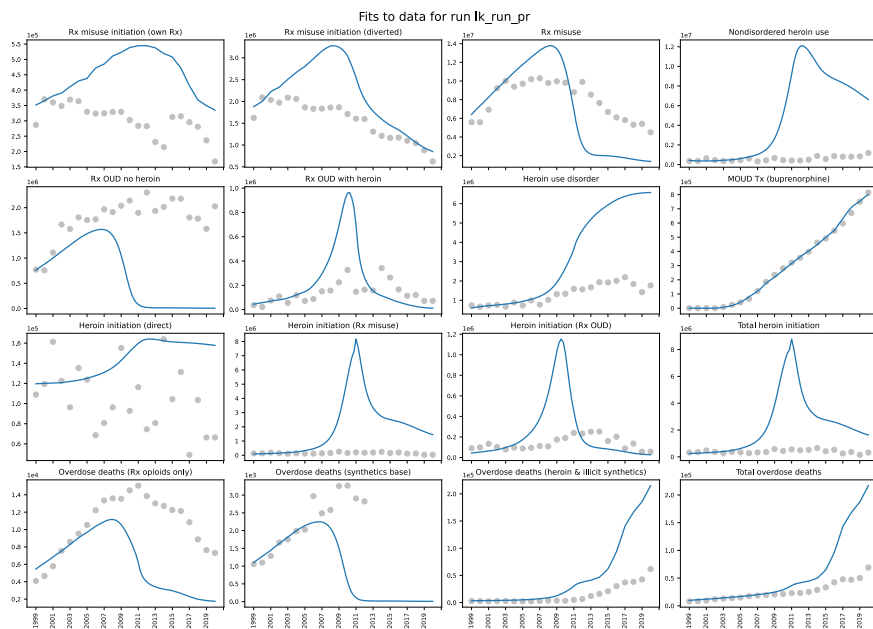
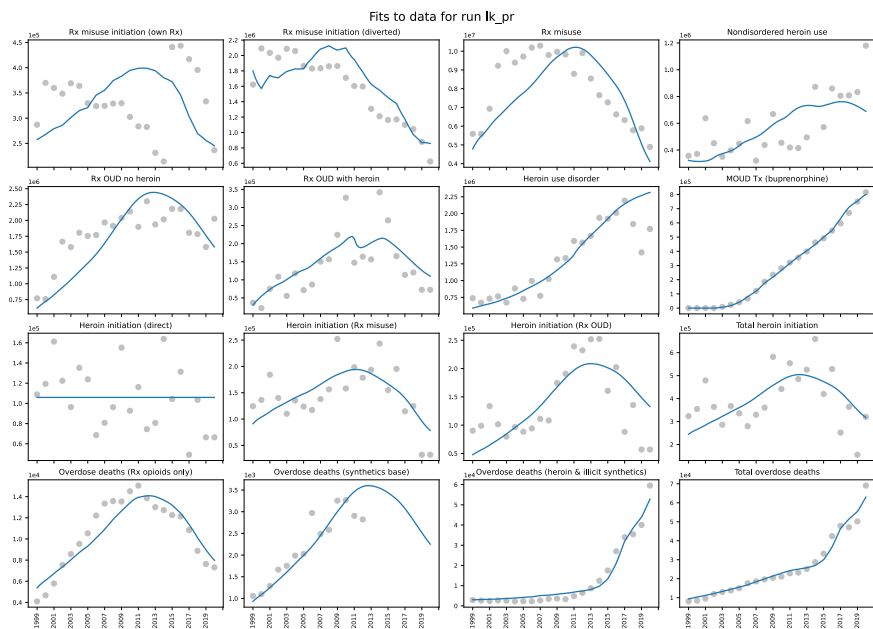
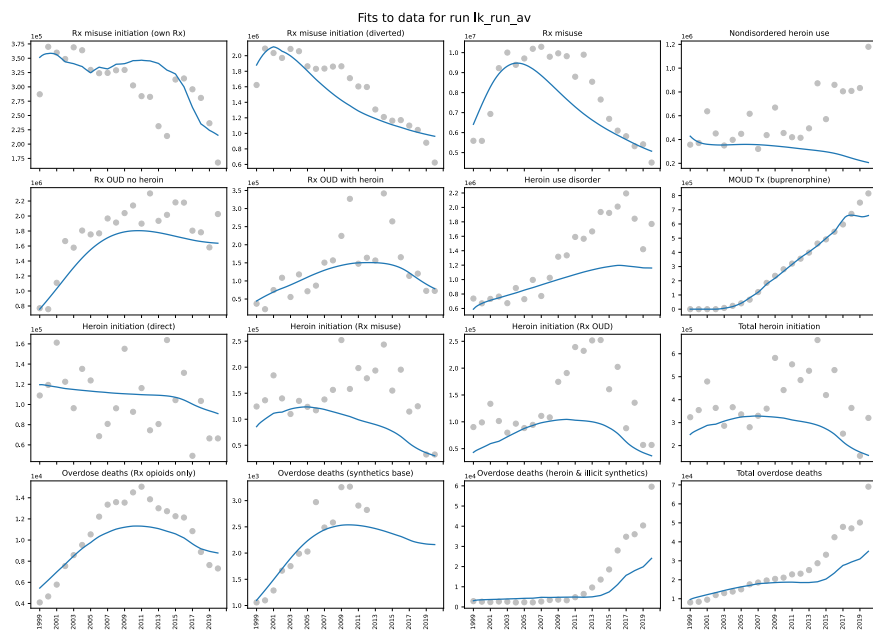
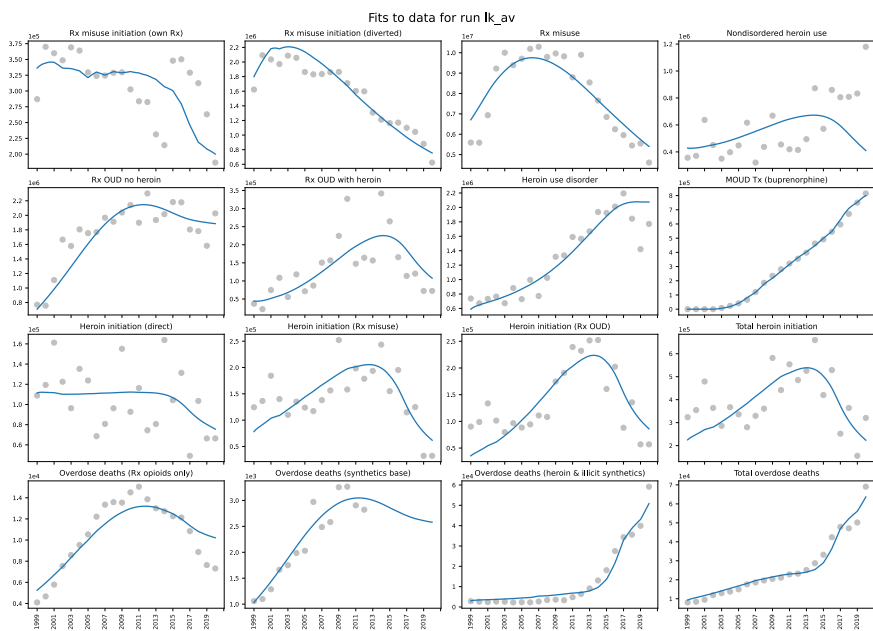
In particular, social influence and availability effects appear to operate similarly, and are able to compensate for each other to a large extent, albeit more so for social influence compensating for availability effects than vice-versa. Social influence feedbacks play a greater role in driving declines in both Rx and heroin initiation in the last few years and going forward. Overall, the similarity in effects of the two processes is not entirely surprising – in practice, access to drugs is not only a matter of aggregate forces of supply and demand but also ease of access within particular social networks (see S2.c.i)), e.g., whether you have a friend who knows a dealer. While we have operationalised our availability construct to reflect aggregate market forces, there likely remains some overlap between social influence and availability feedback processes.

Consistent with the ‘deactivated’ tests, the perceived risk feedbacks seem to play the greatest role, with the other feedback processes less able to compensate for their absence. Without the risk response, rates of both Rx and heroin initiation are unable to decline as much or as quickly as observed in the data, supporting the potential importance of risk responses in producing the ‘wave’ pattern observed in past substance use crises.

Interestingly, all three ‘recalibrated’ cases show increased projected deaths and use disorder, indicating that at this point, and over the next decade, all three sets of feedbacks are likely working in the direction of lessening the magnitude of the crisis. This observation provides some justification for the optimism of our base case projections (S5.c)), which show that the crisis, in terms of use disorder and mortality, appears to be on the verge of peaking and turning around.

Table S15. Sensitivity of key outcomes to loop knockout analyses. ‘Deactivated’ results are from model runs using baseline estimated values with the focal loop[s] subsequently deactivated; ‘Recalibrated’ results deactivate the focal loop[s] and re-estimate the model, using the re-estimated parameters for projections as well.

	Cumulative overdose deaths	Cumulative UD person-years	Projected cumulative overdose deaths	Projected cumulative UD person-years	Med elasticity	Max elasticity	Avg MAEN	Max MAEN
Deactivated Availability	-0.251	-0.170	-0.401	-0.156	0	0	0.228	0.535
Recalibrated w/o Availability	-0.001	-0.003	0.062	0.143	0.045	2.6	0.140	0.329
Deactivated Perceived risk	1.121	0.269	3.081	0.919	0	0	1.972	12.376
Recalibrated w/o Perceived risk	-0.005	0.004	0.126	0.007	0.059	18.6	0.155	0.301
Deactivated Social influence	-0.139	-0.045	-0.320	-0.092	0	0	0.237	0.505
Recalibrated w/o Social influence	-0.004	-0.001	0.116	-0.034	0.072	2.5	0.136	0.294



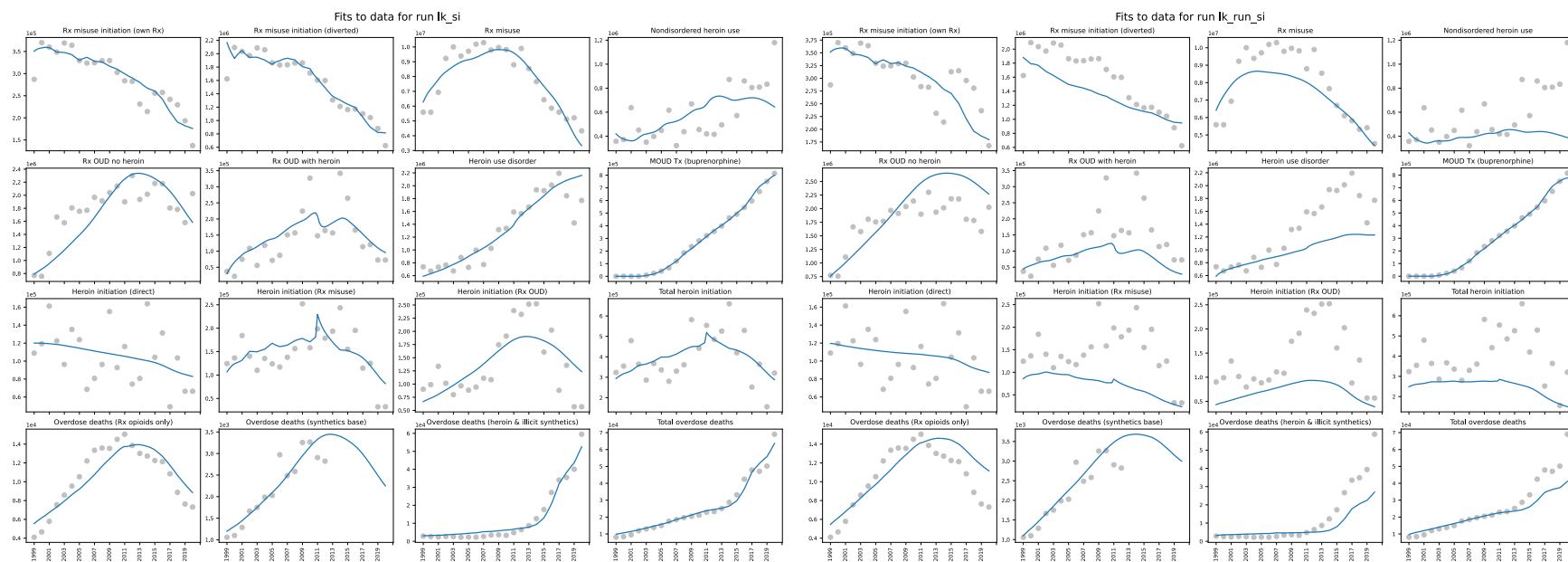


Figure S20. Model fit to data for loop knockout analyses, for all time-series data used in estimation (simulated model output = blue, historical data = grey). Left panels show results of recalibration with loop knockout; right panels show impact of knockout with baseline parameter estimates. Focal loops deactivated, from top to bottom: availability effects, perceived risk, social influence.

S6.c) Sensitivity of projections to exogenous input assumptions

SOURCE makes projections of potential future trajectories of the opioid crisis using some baseline assumptions about future trends in exogenous inputs x_{jt} . By default, we include three main sets of baseline assumptions about SOURCE's inputs (see **Table S13**): an 'exogenous trends continue' (ETC) case where present trends continue at decelerating rates, stabilising at plausible levels by 2032, as well as 'optimistic' and 'pessimistic' cases (see S5.c)). Alternative sets of baseline assumptions can be specified by model users.

To quantify sensitivity to baseline assumptions, we also test a 'constant' case where exogenous inputs do not change after their last data points in 2020. Different baseline assumptions for inputs will obviously have substantial effects on parts of the model directly driven by those inputs. For instance, switching between ETC and 'constant' assumptions about future opioid prescribing trends results in large differences in prescription opioid supply. However, major downstream outcomes are not very sensitive to the baseline case used. Switching from ETC to 'constant' assumptions changes total use disorder prevalence (measured in cumulative person-years from 2020-2032) by 8.6%, and total cumulative overdose deaths by -7.1%.

To assess the impacts of baseline assumptions about individual exogenous inputs on these downstream outcomes, we test each input variable j in two ways: 1) setting all input assumptions except j to their ETC trajectories but holding j constant after 2020; and 2) holding all input assumptions constant after 2020 but setting j to its ETC trajectory. We then calculate the percentage change in key outcomes (cumulative overdose deaths and person-years of use disorder from 2020-2032) from the all-inputs ETC case and constant case, respectively, as well as the mean absolute percentage change for each input j and across all inputs. With the exception of fentanyl penetration in the illicit drug supply, these outcomes are not very sensitive to changes in input assumptions (see **Table S16**) – on average, switching each input between ETC and 'constant' assumptions individually results in changes of 0.0% and 0.0% in projected cumulative use disorder prevalence and overdose deaths respectively.

Table S16. Sensitivity of key projected outcomes to alternative base case assumptions

	Base Projected cumulative overdose deaths	Base Projected cumulative UD person-years	Constant Projected cumulative overdose deaths	Constant Projected cumulative UD person-years	Average Projected cumulative overdose deaths	Average Projected cumulative UD person-years
Fent	-0.19706	0.015097	0.244989	-0.01589	-0.22102	0.015491
NxKD	0.043895	-0.00319	-0.04074	0.002592	0.042316	-0.00289
HPI	0.000369	6.01E-05	-0.00034	-6.80E-05	0.000354	6.40E-05
BMDCap	0.001204	0.000282	-0.00139	-0.0004	0.001297	0.00034
MMTCap	0.070109	0.013703	-0.06216	-0.01336	0.066132	0.013529
VivCap	0.003535	0.00202	-0.00345	-0.00189	0.003493	0.001955
PtRx	0.021198	0.030998	-0.02607	-0.03588	0.023633	0.033436
RxPP	0.002668	0.00386	-0.00445	-0.00637	0.00356	0.005116
RxDur	-0.00436	-0.00604	0.00833	0.011625	-0.00635	-0.00883
MME	0.009216	0.021112	-0.01383	-0.0273	0.011525	0.024204
ADF	-6.79E-05	-5.47E-06	6.14E-05	0	-6.46E-05	-2.73E-06
MAC	0.032153	0.00876	0.036891	0.010487	0.034522	0.009623

For comparison, **Figure S21** shows projections for all time-series data used in model estimation under constant-input assumptions (see **Table S13**) (see **Table S8** in S4.b)). Two main sets of differences stand out. First, deaths fall more sharply than in the ETC case, as fentanyl prevalence does not continue to rise. Second, without continued declines in opioid prescribing, Rx opioid misuse, use disorder, and associated outcomes start to increase again within a few years, as the perceptions of risk associated with Rx misuse that drove its initial decline in the 2000s start to fade – highlighting the importance of continuing various ongoing efforts to reduce prescribing.

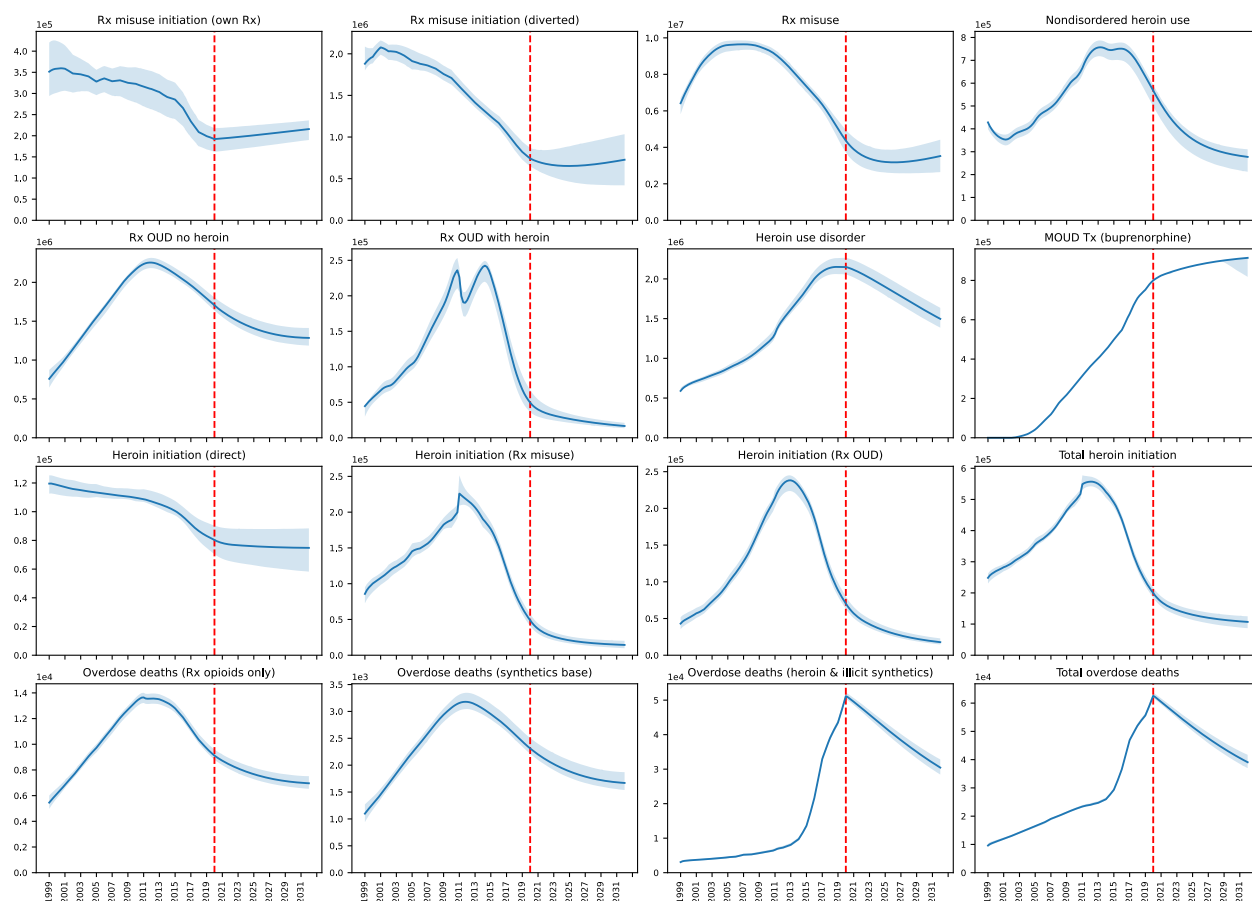


Figure S21. Constant-assumptions case projections for all time-series data used in estimation. Shaded areas indicate 95% credible intervals, not accounting for measurement noise.

S6.d) Sensitivity of parameter estimates to inclusion of 2020 data

As a system dynamics model, SOURCE seeks to reflect the key structural relationships underlying the trajectory of the opioid crisis. Potentially relevant relationships that are not captured within the model, such as the dynamics of heroin and illicit fentanyl supply, are reflected instead in the model's exogenous time-series inputs (see **Table S8**).

The onset of the COVID-19 pandemic in early 2020 thus presents a challenge for the model estimation process. COVID-19 and the resultant socioeconomic impacts are the sort of true exogenous shock unforeseeable in most models. The pandemic has plausibly affected many aspects of the opioid crisis,

such as access to treatment and harm reduction services, risk of relapse to substance use disorder, and initiation or escalation of substance use. Insofar as these COVID-19 impacts are reflected in SOURCE's exogenous inputs (e.g., reduced treatment capacity), the model should accurately account for their broader effects on the crisis. However, if those impacts are *not* reflected in exogenous inputs (e.g., heightened risk of relapse), the model does not structurally account for the resultant effects. Incorporating data from 2020 into the model estimation process thus risks confounding underlying structural relationships with the exogenous impacts of COVID-19, potentially biasing parameter estimates.

One approach to addressing this confounding would be to estimate year fixed effects on initiation, relapse, etc. for 2020 (and any subsequent years under pandemic conditions). However, this approach would entail estimating several additional parameters based on, in most cases, a single data point for each. Instead, here we seek to quantify the magnitude of the confounding problem, in order to demonstrate that the model's exogenous inputs already capture most of the impact of COVID-19 and the extent of remaining confounding is minimal.

To test this, we estimate the model in accordance with S4.c) but using time-series input and calibration target data only through 2019, i.e., pre-pandemic. We then compare resultant parameter estimates and other outcomes as outlined in S6.a) and S6.b).

Outcomes are reported in **Table S17**. In summary, very little changes; the median change in estimated parameter values is ~1%, and projected cumulative overdoses fall by only ~6%. The robustness of estimates and projections indicates that much of the exogenous impact of COVID-19 in 2020 is reflected well in the model's exogenous inputs. While other impacts may have occurred that the model does not account for (e.g., increased relapse rates), those shocks have little effect on the model's long-term behaviour.

S7) Model Equations

A full listing of model equations is available online at <https://github.com/FDA/SOURCE>.

Table S17. Sensitivity of key outcomes to exclusion of 2020 data

	Cumulative overdose deaths	Cumulative UD person- years	Projected cumulative overdose deaths	Projected cumulative UD person- years	Med elasticity	Max elasticity	Avg MAEN	Max MAEN
Excluding 2020 data	-0.00788	0.006924	0.019055	0.102588	0.009327	0.500095	0.123994	0.251937

S8) References

1. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, 5th Ed. (American Psychiatric Association, 2013) <https://doi.org/https://doi.org/10.1176/appi.books.9780890425596>.
2. Center for Behavioral Health Statistics and Quality, "National Survey on Drug Use and Health: 2014 and 2015 redesign changes" (2015).
3. D. Fast, W. Small, E. Wood, T. Kerr, Coming "down here": young people's reflections on becoming entrenched in a local drug scene. *Soc. Sci. Med.* **69**, 1204–10 (2009).
4. H. Guarino, P. Mateu-Gelabert, J. Teubl, E. Goodbody, Young adults' opioid use trajectories: From nonmedical prescription opioid use to heroin, drug injection, drug treatment and overdose. *Addict. Behav.* **86**, 118–123 (2018).
5. A. Harocopos, B. Allen, D. Paone, Circumstances and contexts of heroin initiation following non-medical opioid analgesic use in New York City. *Int. J. Drug Policy* **28**, 106–112 (2016).
6. A. M. Lovell, Risking risk: The influence of types of capital and social networks on the injection practices of drug users. *Soc. Sci. Med.* **55**, 803–821 (2002).
7. S. G. Mars, P. Bourgois, G. Karandinos, F. Montero, D. Ciccarone, "Every 'never' I ever said came true": Transitions from opioid pills to heroin injecting. *Int. J. Drug Policy* **25**, 257–266 (2014).
8. É. Roy, É. Nonn, N. Haley, Transition to injection drug use among street youth-A qualitative analysis. *Drug Alcohol Depend.* **94**, 19–29 (2008).
9. K. Debeck, *et al.*, A dose-dependent relationship between exposure to a street-based drug scene and health-related harms among people who use injection drugs. *J. Urban Heal.* **88**, 724–735 (2011).
10. D. F. Musto, *The American Disease: Origins of narcotic control*, 3rd Ed. (Oxford University Press, 1999).
11. C. P. Rydell, J. P. Caulkins, S. S. Everingham, Enforcement or Treatment? Modeling the Relative Efficacy of Alternatives for Controlling Cocaine. *Oper. Res.* **44**, 687–695 (1996).
12. A. M. Arria, K. M. Caldeira, K. B. Vincent, K. E. O'Grady, E. D. Wish, Perceived harmfulness predicts nonmedical use of prescription drugs among college students: Interactions with sensation-seeking. *Prev. Sci.* **9**, 191–201 (2008).
13. G. S. Becker, Crime and Punishment: An Economic Approach. *J. Polit. Econ.* **76**, 169–217 (1968).
14. D. S. Nagin, R. Paternoster, The preventive effects of the perceived risk of arrest: Testing an expanded conception of deterrence. *Criminology* **29**, 561–587 (1991).
15. A. M. Gill, R. J. Michaels, The Determinants of Illegal Drug Use. *Contemp. Econ. Policy* **9**, 93–105 (1991).
16. S. R. Friedman, *et al.*, Drug arrests and injection drug deterrence. *Am. J. Public Health* **101**, 344–349 (2011).
17. R. J. MacCoun, Drugs and the law: A psychological analysis of drug prohibition. *Psychol. Bull.* **113**, 497–512 (1993).
18. R. Moorman-Li, *et al.*, A review of abuse-deterrent opioids for chronic nonmalignant pain. *Pharm. Ther.* **37**, 412–421 (2012).
19. S. E. McCabe, J. A. Cranford, C. J. Boyd, C. J. Teter, Motives, diversion and routes of administration associated with nonmedical use of prescription opioids. *Addict. Behav.* **32**, 562–575 (2007).
20. R. G. Carlson, R. W. Nahhas, S. S. Martins, R. Daniulaityte, Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study. *Drug Alcohol*

- Depend.* **160**, 127–134 (2016).
21. J. D. Sterman, *Business Dynamics* (McGraw-Hill, 2000).
 22. Y. Kaya, K. Yokobori, Eds., *Environment, energy, and economy: Strategies for sustainability* (United Nations University Press, 1997).
 23. K. T. Vo, X. M. R. van Wijk, K. L. Lynch, A. H. B. Wu, C. G. Smollin, Counterfeit Norco Poisoning Outbreak - San Francisco Bay Area, California, March 25-April 5, 2016. *Morb. Mortal. Wkly. Rep.* **65**, 420–423 (2016).
 24. A. M. Arens, *et al.*, Adverse effects from counterfeit alprazolam tablets. *JAMA Intern. Med.* **176**, 1554–1555 (2016).
 25. T. C. Green, M. Gilbert, Counterfeit Medications and Fentanyl. *JAMA Intern. Med.* **176**, 1555 (2016).
 26. R. M. Gladden, J. O'Donnell, C. L. Mattson, P. Seth, Changes in Opioid-Involved Overdose Deaths by Opioid Type and Presence of Benzodiazepines, Cocaine, and Methamphetamine - 25 States, July-December 2017 to January-June 2018. *Morb. Mortal. Wkly. Rep.* **68**, 737–744 (2019).
 27. J. K. O'Donnell, J. Halpin, C. L. Mattson, B. A. Goldberger, R. M. Gladden, Deaths involving fentanyl, fentanyl analogs, and U-47700 - 10 states, July-December 2016. *Morb. Mortal. Wkly. Rep.* **66**, 1197–1202 (2017).
 28. Drug Enforcement Administration, “Counterfeit prescription pills containing fentanyls: A global threat” (2016).
 29. Drug Enforcement Administration, “2020 National Drug Threat Assessment” (2021) <https://doi.org/10.36548/jtcsst.2021.1>.
 30. J. Y. Park, L. T. Wu, Sources of Misused Prescription Opioids and Their Association with Prescription Opioid Use Disorder in the United States: Sex and Age Differences. *Subst. Use Misuse* **55**, 928–936 (2020).
 31. B. Han, *et al.*, Prescription opioid use, misuse, and use disorders in U.S. Adults: 2015 National Survey on Drug Use and Health. *Ann. Intern. Med.* **167**, 293–301 (2017).
 32. T. J. Cicero, M. S. Ellis, Z. A. Kasper, Relative preferences in the abuse of immediate-release versus extended-release opioids in a sample of treatment-seeking opioid abusers. *Pharmacoepidemiol. Drug Saf.* **26**, 56–62 (2017).
 33. D. Ciccarone, US Heroin in Transition: Supply Changes, Fentanyl Adulteration and Consequences. *Int. J. Drug Policy* **46**, 107–111 (2017).
 34. J. P. Caulkins, P. Reuter, Illicit drug markets and economic irregularities. *Socioecon. Plann. Sci.* **40**, 1–14 (2006).
 35. M. Bouchard, On the resilience of illegal drug markets. *Glob. Crime* **8**, 325–344 (2007).
 36. Drug Enforcement Administration, “National Heroin Threat Assessment Summary - Updated” (2016) (March 19, 2021).
 37. Drug Enforcement Administration, “The 2016 Heroin Signature Program Report” (2018) (March 19, 2021).
 38. US Congress, Comprehensive Addiction and Recovery Act of 2016 (2016).
 39. A. Duncan, J. Anderman, T. Deseran, I. Reynolds, B. D. Stein, Monthly Patient Volumes of Buprenorphine-Waivered Clinicians in the US. *JAMA Netw. Open* **3**, e2014045 (2020).
 40. C. P. Thomas, *et al.*, Prescribing patterns of buprenorphine waived physicians. *Drug Alcohol Depend.* **181**, 213–218 (2017).
 41. K. Mackey, S. Veazie, J. Anderson, D. Bourne, K. Peterson, Barriers and Facilitators to the Use of Medications

- for Opioid Use Disorder: a Rapid Review. *J. Gen. Intern. Med.* **35**, 954–963 (2020).
42. M. Fatseas, M. Auriacombe, Why buprenorphine is so successful in treating opiate addiction in France. *Curr. Psychiatry Rep.* **9**, 358–364 (2007).
 43. J. R. Langabeer, *et al.*, Disparities between US Opioid Overdose Deaths and Treatment Capacity: A Geospatial and Descriptive Analysis. *J. Addict. Med.* **13**, 476–482 (2019).
 44. C. W. Jones, *et al.*, Comparison between buprenorphine provider availability and opioid deaths among US counties. *J. Subst. Abuse Treat.* **93**, 19–25 (2018).
 45. US Congress, Drug Addiction Treatment Act of 2000 (2000).
 46. US Congress, Drug Addiction Treatment Expansion Act (2005).
 47. Drug Enforcement Administration Diversion Control Division, “NFLIS Brief: Fentanyl, 2001-2015” (2017).
 48. Drug Enforcement Administration Diversion Control Division, “NFLIS Brief: Fentanyl and Fentanyl-Related Substances Reported in NFLIS, 2015-2016” (2017).
 49. J. Suzuki, S. El-Haddad, A review: Fentanyl and non-pharmaceutical fentanyls. *Drug Alcohol Depend.* **171**, 107–116 (2017).
 50. T. J. Cicero, M. S. Ellis, H. L. Surratt, S. P. Kurtz, The changing face of heroin use in the United States: A retrospective analysis of the past 50 years. *JAMA Psychiatry* **71**, 821–826 (2014).
 51. D. Ciccarone, The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis. *Int. J. Drug Policy* **71**, 183–188 (2019).
 52. Center for Behavioral Health Statistics and Quality, “Impact of the DSM-IV to DSM-5 changes on the National Survey on Drug Use and Health” (2016).
 53. P. Reuter, J. P. Caulkins, G. Midgette, Heroin use cannot be measured adequately with a general population survey. *Addiction* **116**, 2600–2609 (2021).
 54. J. A. Barocas, *et al.*, Estimated Prevalence of Opioid Use Disorder in Massachusetts, 2011-2015: A Capture-Recapture Analysis. *Am. J. Public Health* **108**, 1675–1681 (2018).
 55. W. M. Compton, D. Dawson, S. Q. Duffy, B. F. Grant, The effect of inmate populations on estimates of DSM-IV alcohol and drug use disorders in the United States. *Am. J. Psychiatry* **167**, 473–474 (2010).
 56. G. Midgette, S. Davenport, J. P. Caulkins, B. Kilmer, “What America’s Users Spend on Illegal Drugs, 2006-2016” (2019).
 57. R. A. Grucza, A. M. Abbacchi, T. R. Przybeck, J. C. Gfroerer, Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys. *Addiction* **102**, 623–629 (2007).
 58. W. M. Compton, D. a. Dawson, R. B. Goldstein, B. F. Grant, Crosswalk between DSM-IV dependence and DSM-5 substance use disorders for opioids, cannabis, cocaine and alcohol. *Drug Alcohol Depend.* **132**, 387–390 (2013).
 59. B. Kilmer, *et al.*, “What America’s Users Spend on Illegal Drugs: 2000-2010” (2014).
 60. N. Dasgupta, *et al.*, Inches, Centimeters, and Yards: Overlooked definition choices inhibit interpretation of morphine equivalence. *Clin. J. Pain Publish Ah* (2021).
 61. U.S. Food and Drug Administration, Morphine Milligram Equivalents: Current Applications and Knowledge Gaps, Research Opportunities, and Future Directions (2021) (July 8, 2021).
 62. UNODC, Heroin and cocaine prices in Europe and USA | dataUNODC. *World Drug Rep.* (2018).

63. J. Homer, W. Wakeland, A dynamic model of the opioid drug epidemic with implications for policy. *Am. J. Drug Alcohol Abuse* (2020) <https://doi.org/10.1080/00952990.2020.1755677>.
64. Substance Abuse and Mental Health Services Administration, 2018-2019 National Survey on Drug Use and Health: Model-Based Prevalence Estimates (50 States and the District of Columbia) (2019).
65. J. Gryczynski, *et al.*, Patient perspectives on choosing buprenorphine over methadone in an urban, equal-access system. *Am. J. Addict.* **22**, 285–291 (2013).
66. B. J. H. Yarborough, *et al.*, Methadone, buprenorphine and preferences for opioid agonist treatment: A qualitative analysis. *Drug Alcohol Depend.* **160**, 112–118 (2016).
67. T. Beetham, B. Saloner, S. E. Wakeman, M. Gaye, M. L. Barnett, Access to office-based buprenorphine treatment in areas with high rates of opioid-related mortality: An audit study. *Ann. Intern. Med.* **171**, 1–9 (2019).
68. X. Wan, W. Wang, J. Liu, T. Tong, Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med. Res. Methodol.* **14**, 1–13 (2014).
69. H. S. Connery, R. D. Weiss, Discontinuing buprenorphine treatment of opioid use disorder: What do we (not) know? *Am. J. Psychiatry* **177**, 104–106 (2020).
70. B. Eastwood, J. Strang, J. Marsden, Effectiveness of treatment for opioid use disorder: A national, five-year, prospective, observational study in England. *Drug Alcohol Depend.* **176**, 139–147 (2017).
71. J. S. Potter, *et al.*, The multi-site prescription opioid addiction treatment study: 18-month outcomes. *J. Subst. Abuse Treat.* **48**, 62–69 (2015).
72. M.-J. Fleury, *et al.*, Remission from substance use disorders: A systematic review and meta-analysis. *Drug Alcohol Depend.* **168**, 293–306 (2016).
73. J. F. Kelly, B. Bergman, B. B. Hoepfner, C. Vilsaint, W. L. White, Prevalence and pathways of recovery from drug and alcohol problems in the United States population: Implications for practice, research, and policy. *Drug Alcohol Depend.* **181**, 162–169 (2017).
74. Y. I. Hser, E. Evans, C. Grella, W. Ling, D. Anglin, Long-term course of opioid addiction. *Harv. Rev. Psychiatry* **23**, 76–89 (2015).
75. R. L. DuPont, W. M. Compton, A. T. McLellan, Five-Year Recovery: A New Standard for Assessing Effectiveness of Substance Use Disorder Treatment. *J. Subst. Abuse Treat.* **58**, 1–5 (2015).
76. National Center for Health Statistics, Mortality Multiple Cause-of-Death Public Use Data Files and Documentation (2021) (March 21, 2021).
77. N. Wilson, M. Kariisa, P. Seth, H. Smith IV, N. L. Davis, Drug and Opioid-Involved Overdose Deaths — United States, 2017–2018. *Morb. Mortal. Wkly. Rep.* **69**, 290–297 (2020).
78. Drug Enforcement Administration Office of Diversion Control, “NFLIS Special Report: Fentanyl, 2003-2006” (2008) <https://doi.org/10.4135/9781544377230.n112>.
79. D. A. Algren, *et al.*, Fentanyl-associated Fatalities Among Illicit Drug Users in Wayne County, Michigan (July 2005-May 2006). *J. Med. Toxicol.* **9**, 106–115 (2013).
80. R. Mrvos, *et al.*, Whole fentanyl patch ingestion: A multi-center case series. *J. Emerg. Med.* **42**, 549–552 (2012).
81. M. I. Jumbelic, Deaths With Transdermal Fentanyl Patches. *Am. J. Forensic Med. Pathol.* **31**, 18–21 (2010).
82. J. J. Carroll, B. D. L. Marshall, J. D. Rich, T. C. Green, Exposure to fentanyl-contaminated heroin and overdose

- risk among illicit opioid users in Rhode Island: A mixed methods study. *Int. J. Drug Policy* **46**, 136–145 (2017).
83. A. Macmadu, J. J. Carroll, S. E. Hadland, T. C. Green, B. D. L. Marshall, Prevalence and correlates of fentanyl-contaminated heroin exposure among young adults who use prescription opioids non-medically. *Addict. Behav.* **68**, 35–38 (2017).
 84. S. G. Sherman, *et al.*, Acceptability of implementing community-based drug checking services for people who use drugs in three United States cities: Baltimore, Boston and Providence. *Int. J. Drug Policy* **68**, 46–53 (2019).
 85. M. C. Meacham, *et al.*, Addressing overdose risk among unstably housed women in San Francisco, California: An examination of potential fentanyl contamination of multiple substances. *Harm Reduct. J.* **17** (2020).
 86. L. Scholl, P. Seth, M. Kariisa, N. Wilson, G. Baldwin, “Drug and Opioid-Involved Overdose Deaths — United States, 2013–2017” (2019).
 87. J. N. Park, B. W. Weir, S. T. Allen, P. Chaulk, S. G. Sherman, Fentanyl-contaminated drugs and non-fatal overdose among people who inject drugs in Baltimore, MD. *Harm Reduct. J.* **15**, 1–8 (2018).
 88. E. Wheeler, T. S. Jones, M. K. Gilbert, P. J. Davidson, Opioid Overdose Prevention Programs Providing Naloxone to Laypersons - United States, 2014. *Morb. Mortal. Wkly. Rep.* **64**, 631–635 (2015).
 89. E. Wheeler, *et al.*, Community-Based Opioid Overdose Prevention Programs Providing Naloxone — United States, 2010. *Morb. Mortal. Wkly. Rep.* **61**, 101–105 (2012).
 90. E. Wheeler, M. Doe-Simkins, Harm Reduction programs distribute one million doses of naloxone in 2019 | by Eliza Wheeler | Medium. *Medium* (2020) (August 31, 2020).
 91. D. Belz, J. Lieb, T. Rea, M. S. Eisenberg, Naloxone use in a tiered-response emergency medical services system. *Prehospital Emerg. Care* **10**, 468–471 (2006).
 92. M. Irvine, *et al.*, Modeling the impact of naloxone distribution for overdose prevention through community programs, prescriptions, and pharmacy-facilitated channels in the US: Results from a 10-state analysis in *Addiction Science & Clinical Practice*, (2020), p. A98.
 93. S. Koester, S. R. Mueller, L. Raville, S. Langeegger, I. A. Binswanger, Why are some people who have received overdose education and naloxone reticent to call Emergency Medical Services in the event of overdose? *Int. J. Drug Policy* **48**, 115–124 (2017).
 94. I. J. Martinez-Moyano, G. P. Richardson, Best practices in system dynamics modeling. *Syst. Dyn. Rev.* **29**, 102–123 (2013).
 95. A. L. Pitt, K. Humphreys, M. L. Brandeau, Modeling health benefits and harms of public policy responses to the US opioid epidemic. *Am. J. Public Health* **108**, 1394–1400 (2018).
 96. W. Wakeland, A. Nielsen, P. Geissert, Dynamic model of nonmedical opioid use trajectories and potential policy interventions. *Am. J. Drug Alcohol Abuse* **41**, 508–518 (2015).
 97. J. Struben, J. D. Sterman, D. Keith, “Parameter estimation through maximum likelihood and bootstrapping methods” in *Analytical Methods for Dynamic Modelers*, 1st Ed., H. Rahmandad, R. Oliva, N. D. Osgood, Eds. (MIT Press, 2015), pp. 3–38.
 98. S. Basu, J. Andrews, Complexity in Mathematical Models of Public Health Policies: A Guide for Consumers of Models. *PLoS Med.* **10**, 1–6 (2013).
 99. J. A. Vrugt, *et al.*, Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Int. J. Nonlinear Sci. Numer. Simul.* **10**, 271–288 (2009).
 100. N. D. Osgood, J. Liu, “Combining Markov Chain Monte Carlo approaches and dynamic modeling” in *Analytical Methods for Dynamic Modelers*, H. Rahmandad, R. Oliva, N. D. Osgood, Eds. (MIT Press, 2015), pp. 125–170.

101. J. B. Homer, Partial-model testing as a validation tool for system dynamics (1983). *Syst. Dyn. Rev.* **28**, 281–294 (2012).
102. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–511 (1992).
103. T. Li, H. Rahmandad, J. D. Sterman, “Improving Parameter Estimation of Epidemic Models: Likelihood Functions and Kalman Filtering” (2021).
104. R. Oliva, Model calibration as a testing strategy for system dynamics models. *Eur. J. Oper. Res.* **151**, 552–568 (2003).
105. H. Theil, *Applied Economic Forecasting* (North Holland Publishing Company, 1966).
106. D. N. Ford, A behavioral approach to feedback loop dominance analysis. *Syst. Dyn. Rev.* **15**, 3–36 (1999).
107. Centers for Disease Control and Prevention, “2019 CDC Opioid NDC and Oral MME Conversion File” (2019).
108. M. L. McPherson, *Demystifying opioid conversion calculations: A guide for effective dosing*, 2nd Ed. (American Society of Health-System Pharmacists, 2018).
109. S. Kishner, E. D. Schraga, Opioid equivalents and conversions. *Medscape* (2018) (February 13, 2019).
110. GlobalRPh, Opioid conversions calc (single agent) equianalgesic (2019) (February 13, 2019).
111. S. L. Proctor, A. L. Copeland, A. M. Kopak, P. L. Herschman, N. Polukhina, A naturalistic comparison of the effectiveness of methadone and two sublingual formulations of buprenorphine on maintenance treatment outcomes: Findings from a retrospective multisite study. *Exp. Clin. Psychopharmacol.* **22**, 424–433 (2014).
112. I. L. Mintzer, *et al.*, Treating Opioid Addiction With Buprenorphine-Naloxone in Community-Based Primary Care Settings. *Ann. Fam. Med.* **5**, 146–150 (2007).
113. A. Khemiri, E. Kharitonova, V. Zah, J. Ruby, M. Toumi, Analysis of buprenorphine/naloxone dosing impact on treatment duration, resource use and costs in the treatment of opioid-dependent adults: A retrospective study of US public and private health care claims. *Postgrad. Med.* **126**, 113–120 (2014).
114. M. Tierney, *et al.*, Two Different Buprenorphine Treatment Settings With Similar Retention Rates: Implications for Expanding Access to Treatment for Opioid Use Disorder. *J. Am. Psychiatr. Nurses Assoc.* **25** (2019).
115. J. D. Lee, *et al.*, Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X:BOT): a multicentre, open-label, randomised controlled trial. *Lancet* **391**, 309–318 (2018).
116. R. E. Clark, *et al.*, Risk Factors for Relapse and Higher Costs Among Medicaid Members with Opioid Dependence or Abuse: Opioid Agonists, Comorbidities, and Treatment History. *J. Subst. Abuse Treat.* **57**, 75–80 (2015).
117. W. H. Lo-Ciganic, *et al.*, Association between trajectories of buprenorphine treatment and emergency department and in-patient utilization. *Addiction* **111**, 892–902 (2016).
118. J. Bell, “Politics, practice and research into treatment of heroin addiction” in *Pharmacotherapies for the Treatment of Opioid Dependence*, (2009) <https://doi.org/10.3109/9780203414088-11>.
119. B. Saloner, *et al.*, A Public Health Strategy for the Opioid Crisis. *Public Health Rep.* **133**, 24S–34S (2018).
120. Y. I. Hser, *et al.*, Long-term outcomes after randomization to buprenorphine/naloxone versus methadone in a multi-site trial. *Addiction* **111**, 695–705 (2016).

121. E. P. Bhatraju, *et al.*, Public sector low threshold office-based buprenorphine treatment: outcomes at year 7. *Addict. Sci. Clin. Pract.* **12**, 7 (2017).
122. H. Samples, A. R. Williams, M. Olfson, S. Crystal, Risk factors for discontinuation of buprenorphine treatment for opioid use disorders in a multi-state sample of Medicaid enrollees. *J. Subst. Abuse Treat.* **95**, 9–17 (2018).
123. E. Clay, *et al.*, Persistence and healthcare utilization associated with the use of buprenorphine/naloxone film and tablet formulation therapy in adults with opioid dependence. *J. Med. Econ.* **17**, 626–636 (2014).
124. A. Manhapra, E. Agbese, D. L. Leslie, R. A. Rosenheck, Three-year retention in buprenorphine treatment for opioid use disorder among privately insured adults. *Psychiatr. Serv.* **69**, 768–776 (2018).
125. J. Bell, C. Mutch, Treatment retention in adolescent patients treated with methadone or buprenorphine for opioid dependence: A file review. *Drug Alcohol Rev.* **25**, 167–171 (2006).
126. M. Hickman, *et al.*, The impact of buprenorphine and methadone on mortality: a primary care cohort study in the United Kingdom. *Addiction* **113**, 1461–1476 (2018).
127. B. Saloner, M. Daubresse, G. C. Alexander, Patterns of Buprenorphine-Naloxone Treatment for Opioid Use Disorder in a Multistate Population. *Med. Care* **55**, 669–676 (2017).
128. J. Gryczynski, *et al.*, Fentanyl exposure and preferences among individuals starting treatment for opioid use disorder. *Drug Alcohol Depend.* **204**, 107515 (2019).
129. A. G. Robertson, *et al.*, Associations between pharmacotherapy for opioid dependence and clinical and criminal justice outcomes among adults with co-occurring serious mental illness. *J. Subst. Abuse Treat.* **86**, 17–25 (2018).
130. J. R. Morgan, B. R. Schackman, J. A. Leff, B. P. Linas, A. Y. Walley, Injectable naltrexone, oral naltrexone, and buprenorphine utilization and discontinuation among individuals treated for opioid use disorder in a United States commercially insured population. *J. Subst. Abuse Treat.* **85**, 90–96 (2018).
131. N. Shcherbakova, G. Tereso, J. Spain, R. J. Roose, Treatment Persistence Among Insured Patients Newly Starting Buprenorphine/Naloxone for Opioid Use Disorder. *Ann. Pharmacother.* **52**, 405–414 (2018).
132. S. E. Hadland, *et al.*, Receipt of Timely Addiction Treatment and Association of Early Medication Treatment with Retention in Care among Youths with Opioid Use Disorder. *JAMA Pediatr.* **172**, 1029–1037 (2018).
133. S. Amiri, *et al.*, Three-year retention in methadone opioid agonist treatment: A survival analysis of clients by dose, area deprivation, and availability of alcohol and cannabis outlets. *Drug Alcohol Depend.* **193**, 63–68 (2018).
134. S. M. Kelly, K. E. O’Grady, S. G. Mitchell, B. S. Brown, R. P. Schwartz, Predictors of methadone treatment retention from a multi-site study: A survival analysis. *Drug Alcohol Depend.* **117**, 170–175 (2011).
135. H. S. Reisinger, *et al.*, Premature discharge from methadone treatment: Patient perspectives. *J. Psychoactive Drugs* **41**, 285–296 (2009).
136. B. Nosyk, *et al.*, Proportional hazards frailty models for recurrent methadone maintenance treatment. *Am. J. Epidemiol.* **170**, 783–792 (2009).
137. E. C. Strain, G. E. Bigelow, I. A. Liebson, M. L. Stitzer, Moderate- vs high-dose methadone in the treatment of opioid dependence: A randomized trial. *J. Am. Med. Assoc.* **281**, 1000–1005 (1999).
138. A. J. Saxon, *et al.*, Extended-release naltrexone (XR-NTX) for opioid use disorder in clinical practice: Vivitrol’s Cost and Treatment Outcomes Registry. *Addiction* **113**, 1477–1487 (2018).
139. E. Krupitsky, *et al.*, Injectable extended-release naltrexone (XR-NTX) for opioid dependence: long-term safety and effectiveness. *Addiction* **108**, 1628–1637 (2013).

140. G. Chang, *et al.*, Adherence to extended release naltrexone: Patient and treatment characteristics. *Am. J. Addict.* **27**, 524–530 (2018).
141. S. M. Murphy, J. R. Morgan, P. J. Jeng, B. R. Schackman, Will converting naloxone to over-the-counter status increase pharmacy sales? *Health Serv. Res.* **54**, 764–772 (2019).
142. L. Degenhardt, *et al.*, Mortality among regular or dependent users of heroin and other opioids: A systematic review and meta-analysis of cohort studies. *Addiction* **106**, 32–51 (2011).
143. L. Sordo, *et al.*, Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies. *BMJ* **357**, j1550 (2017).
144. E. Kelty, G. Hulse, Fatal and non-fatal opioid overdose in opioid dependent patients treated with methadone, buprenorphine or implant naltrexone. *Int. J. Drug Policy* **46**, 54–60 (2017).
145. E. Kelty, D. Joyce, G. Hulse, A retrospective cohort study of mortality rates in patients with an opioid use disorder treated with implant naltrexone, oral methadone or sublingual buprenorphine. *Am. J. Drug Alcohol Abuse* **45**, 285–291 (2019).
146. A. Bahji, B. Cheng, S. Gray, H. Stuart, Reduction in mortality risk with opioid agonist therapy: a systematic review and meta-analysis. *Acta Psychiatr. Scand.* **140**, 313–339 (2019).
147. J. Ma, *et al.*, Effects of medication-assisted treatment on mortality among opioids users: a systematic review and meta-analysis. *Mol. Psychiatry* **24**, 1868–1883 (2019).
148. B. D. Stein, *et al.*, Where Is Buprenorphine Dispensed to Treat Opioid Use Disorders? The Role of Private Offices, Opioid Treatment Programs, and Substance Abuse Treatment Facilities in Urban and Rural Counties. *Milbank Q.* **93**, 561–583 (2015).
149. C. M. Jones, M. Campopiano, G. Baldwin, E. Mccance-Katz, National and State Treatment Need and Capacity for Opioid Agonist Medication-Assisted Treatment (2015) <https://doi.org/10.2105/AJPH>.
150. US Department of Health and Human Services, Medication Assisted Treatment for Opioid Use Disorder (2016).
151. Congressional Research Service, “Buprenorphine and the Opioid Crisis: A Primer for Congress” (2018).
152. SAMHSA, Practitioner and Program Data (2020).
153. L. Degenhardt, *et al.*, Global patterns of opioid use and dependence: harms to populations, interventions, and future action. *Lancet* **394**, 1560–1579 (2019).
154. J. R. Morgan, B. R. Schackman, Z. M. Weinstein, A. Y. Walley, B. P. Linas, Overdose following initiation of naltrexone and buprenorphine medication treatment for opioid use disorder in a United States commercially insured cohort. *Drug Alcohol Depend.* **200**, 34–39 (2019).
155. M. R. Larochelle, *et al.*, Medication for opioid use disorder after nonfatal opioid overdose and association with mortality: A cohort study. *Ann. Intern. Med.* **169**, 137–145 (2018).
156. L.-T. Wu, G. Woody, C. Yang, P. Mannelli, Blazer, Differences in onset and abuse/dependence episodes between prescription opioids and heroin: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Subst. Abuse Rehabil.*, 77 (2011).
157. W. M. Compton, Y. F. Thomas, F. S. Stinson, B. F. Grant, Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *Arch. Gen. Psychiatry* **64**, 566–76 (2007).
158. C. E. Grella, M. P. Karno, U. S. Warda, N. Niv, A. A. Moore, Gender and comorbidity among individuals with opioid use disorders in the NESARC study. *Addict. Behav.* **34**, 498–504 (2009).

159. L. A. Hoffman, C. Vilsaint, J. F. Kelly, Recovery From Opioid Problems in the US Population. *J. Addict. Med.* **00**, 1 (2019).
160. S. S. Martins, *et al.*, Changes in US Lifetime Heroin Use and Heroin Use Disorder. *JAMA Psychiatry* **74**, 445 (2017).
161. S. Galea, *et al.*, Provision of naloxone to injection drug users as an overdose prevention strategy: Early evidence from a pilot study in New York City. *Addict. Behav.* **31**, 907–912 (2006).
162. J. G. Katzman, *et al.*, Association of Take-Home Naloxone and Opioid Overdose Reversals Performed by Patients in an Opioid Treatment Program. *JAMA Netw. open* **3**, e200117 (2020).
163. M. Doe-Simkins, A. Y. Walley, A. Epstein, P. Moyer, Saved by the nose: Bystander-administered intranasal naloxone hydrochloride for opioid overdose. *Am. J. Public Health* **99**, 788–791 (2009).
164. L. Enteen, *et al.*, Overdose prevention and naloxone prescription for opioid users in san Francisco. *J. Urban Heal.* **87**, 931–941 (2010).
165. K. E. Tobin, S. G. Sherman, P. Beilenson, C. Welsh, C. A. Latkin, Evaluation of the Staying Alive programme: Training injection drug users to properly administer naloxone and save lives. *Int. J. Drug Policy* **20**, 131–136 (2009).
166. J. Strang, *et al.*, Overdose training and take-home naloxone for opiate users: Prospective cohort study of impact on knowledge and attitudes and subsequent management of overdoses. *Addiction* **103**, 1648–1657 (2008).
167. P. Leece, *et al.*, Process evaluation of the Prevent Overdose in Toronto (POINT) program. *Can. J. Public Heal.* **107**, e224–e230 (2016).
168. M. A. Yokell, S. Bowman, W. Alpert, “Opioid Overdose Prevention and Naloxone Distribution in Rhode Island” (2011).
169. O. Banjo, *et al.*, A quantitative and qualitative evaluation of the British Columbia Take Home Naloxone program. *C. Open* **2**, E153–E161 (2014).
170. P. N. Leece, *et al.*, Development and implementation of an opioid overdose prevention and response program in Toronto, Ontario. *Can. J. Public Heal.* **104**, e200–e204 (2013).
171. R. L. Gaston, D. Best, V. Manning, E. Day, Can we prevent drug related deaths by training opioid users to recognise and manage overdoses? *Harm Reduct. J.* **6**, 26 (2009).
172. S. Darke, D. Zador, Fatal heroin “overdose”: A review. *Addiction* **91**, 1765–1772 (1996).
173. C. McGregor, S. Darke, R. Ali, P. Christie, Experience of non-fatal overdose among heroin users in Adelaide, Australia: Circumstances and risk perceptions. *Addiction* **93**, 701–711 (1998).
174. Strang J, *et al.*, Preventing opiate overdose fatalities with take-home naloxone: pre-launch study of possible impact and acceptability. *Addiction* **94**, 199–204 (1999).
175. B. Sergeev, A. Karpets, A. Sarang, M. Tikhonov, “Prevalence and Circumstances of Opiate Overdose Among Injection Drug Users in the Russian Federation” (2003).
176. K. E. Tobin, M. A. Davey, C. A. Latkin, Calling emergency medical services during drug overdose: An examination of individual, social and setting correlates. *Addiction* **100**, 397–404 (2005).
177. A. S. B. Bohnert, M. Tracy, S. Galea, Characteristics of drug users who witness many overdoses: Implications for overdose prevention. *Drug Alcohol Depend.* **120**, 168–173 (2012).
178. R. P. Ogeil, *et al.*, Pharmaceutical opioid overdose deaths and the presence of witnesses. *Int. J. Drug Policy*

- 55, 8–13 (2018).
179. B. Levy, *et al.*, Recognition and response to opioid overdose deaths—New Mexico, 2012. *Drug Alcohol Depend.* **167**, 29–35 (2016).
 180. C. L. Mattson, *et al.*, Opportunities to Prevent Overdose Deaths Involving Prescription and Illicit Opioids, 11 States, July 2016–June 2017. *MMWR. Morb. Mortal. Wkly. Rep.* **67**, 945–951 (2018).
 181. J. O'Donnell, R. M. Gladden, C. L. Mattson, C. T. Hunter, N. L. Davis, “Morbidity and Mortality Weekly Report Vital Signs: Characteristics of Drug Overdose Deaths Involving Opioids and Stimulants-24 States and the District of Columbia” (2020).
 182. N. J. Somerville, *et al.*, Characteristics of fentanyl overdose—Massachusetts, 2014–2016. *MMWR. Morb. Mortal. Wkly. Rep.* **66**, 382 (2017).
 183. P. J. Davidson, K. C. Ochoa, J. A. Hahn, J. L. Evans, A. R. Moss, Witnessing heroin-related overdoses: The experiences of young injectors in San Francisco. *Addiction* **97**, 1511–1516 (2002).
 184. K. H. Seal, *et al.*, Attitudes about Prescribing Take-Home Naloxone to Injection Drug Users for the Management of Heroin Overdose: A Survey of Street-Recruited Injectors in the San Francisco Bay Area. *J. Urban Heal.* **80**, 291–301 (2003).
 185. K. H. Seal, *et al.*, Naloxone distribution and cardiopulmonary resuscitation training for injection drug users to prevent heroin overdose death: A pilot intervention study. *J. Urban Heal.* **82**, 303–311 (2005).
 186. M. Tracy, *et al.*, Circumstances of witnessed drug overdose in New York City: Implications for intervention. *Drug Alcohol Depend.* **79**, 181–190 (2005).
 187. A. S. B. Bohnert, M. Tracy, S. Galea, Circumstances and Witness Characteristics Associated With Overdose Fatality. *Ann. Emerg. Med.* **54**, 618–624 (2009).
 188. R. A. Pollini, *et al.*, Response to Overdose Among Injection Drug Users (2006) <https://doi.org/10.1016/j.amepre.2006.04.002> (December 14, 2020).
 189. S. G. Sherman, *et al.*, A qualitative study of overdose responses among Chicago IDUs. *Harm Reduct. J.* **5** (2008).
 190. A. Y. Walley, *et al.*, Opioid overdose rates and implementation of overdose education and nasal naloxone distribution in Massachusetts: Interrupted time series analysis. *BMJ* **346**, 1–13 (2013).
 191. A. S. Bennett, A. Bell, L. Tomedi, E. G. Hulse, A. H. Kral, Characteristics of an overdose prevention, response, and naloxone distribution program in Pittsburgh and Allegheny County, Pennsylvania. *J. Urban Heal.* **88**, 1020–1030 (2011).
 192. D. Best, *et al.*, Peer overdose resuscitation: multiple intervention strategies and time to response by drug users who witness overdose (2009) <https://doi.org/10.1080/0959523021000002732>.
 193. K. D. Wagner, *et al.*, Evaluation of an overdose prevention and response training programme for injection drug users in the Skid Row area of Los Angeles, CA. *Int. J. Drug Policy* **21**, 186–193 (2010).
 194. M. Doe-Simkins, *et al.*, Overdose rescues by trained and untrained participants and change in opioid use among substance-using participants in overdose education and naloxone distribution programs: A retrospective cohort study. *BMC Public Health* **14** (2014).
 195. J. K. Lim, *et al.*, Factors associated with help seeking by community responders trained in overdose prevention and naloxone administration in Massachusetts. *Drug Alcohol Depend.* **204**, 107531 (2019).
 196. S. E. Lankenau, *et al.*, Injection drug users trained by overdose prevention programs: Responses to witnessed overdoses. *J. Community Health* **38**, 133–141 (2013).

197. C. Rowe, *et al.*, Predictors of participant engagement and naloxone utilization in a community-based naloxone distribution program. *Addiction* **110**, 1301–1310 (2015).
198. K. Dwyer, *et al.*, Opioid education and nasal naloxone rescue kits in the emergency department. *West. J. Emerg. Med.* **16**, 381–384 (2015).
199. K. D. Wagner, *et al.*, Association between non-fatal opioid overdose and encounters with healthcare and criminal justice systems: Identifying opportunities for intervention. *Drug Alcohol Depend.* **153**, 215–220 (2015).
200. S. Schiavon, *et al.*, Medical, psychosocial, and treatment predictors of opioid overdose among high risk opioid users. *Addict. Behav.* **86**, 51–55 (2018).
201. J. Ataiants, *et al.*, Overdose response among trained and untrained women with a history of illicit drug use: a mixed-methods examination. *Drugs Educ. Prev. Policy* (2020) <https://doi.org/10.1080/09687637.2020.1818691> (December 10, 2020).
202. D. G. Schwartz, *et al.*, Layperson reversal of opioid overdose supported by smartphone alert: A prospective observational cohort study. *EClinicalMedicine* **25** (2020).
203. S. Darke, R. P. Mattick, L. Degenhardt, The ratio of non-fatal to fatal heroin overdose. *Addiction* **98**, 1169–1171 (2003).
204. J. Neale, A response to Darke et al., ‘The ratio of non-fatal to fatal heroin overdose.’ *Addiction* **98**, 1171–1171 (2003).
205. T. D. Saha, *et al.*, Nonmedical Prescription Opioid Use and DSM-5 Nonmedical Prescription Opioid Use Disorder in the United States. *J. Clin. Psychiatry* **77**, 772–780 (2016).
206. B. Pardo, *et al.*, “The future of fentanyl and other synthetic opioids” (2019).
207. J. E. Schulenberg, *et al.*, Monitoring the future: College Students & Adults Ages 19 – 55. *Monit. Futur.* **2** (2019).
208. SAMHSA, Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health. *HHS Publ. No. PEP19-5068, NSDUH Ser. H-54* **170**, 51–58 (2019).